Information System Design Methodology Based On PERT/CPM Networking And Optimization Techniques

by

ANINDYA BOSE

University of Pittsburgh

1970

N00014-67-A-0402-0004 NR 049-258

INFORMATION SYSTEM DESIGN METHODOLOGY BASED ON PERT/CPM NETWORKING AND OPTIMIZATION TECHNIQUES

By

Anindya Bose

B.A.(Hons.), University of Calcutta, 1958

M.L.S. University of Pittsburgh, 1965

Submitted to the Faculty in the Graduate School of Library and Information Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

1970

This material was produced for the Government under ONR Contract #N00014-67-A-0402-0004, Identifying #NR 049-258. It may be reproduced and used for governmental purpose without payment of royalties. Reproduction of this material for governmental purpose should include this notice in addition to the copyright notice.

DISSERTATION SERIES

Graduate School of Library and Information Sciences

University of Pittsburgh

Frank B. Sessa, Editor
Elinore Thomas, Assistant Editor

1. Selected Educational Objectives for Public Service Librarians: A Taxonomic Approach
   by Anna C. Hall

2. Analysis of Terminology, Form and Structure of Subject Headings in Arabic Literature and Formulation of Rules for Arabic Subject Headings
   by Muhammad Muhammad Aman

3. Standards for Library Education in Burma, Ceylon, India, and Pakistan
   by Anis Khurshid

4. An Experiment to Determine the Effectiveness of an Interactive Tutorial Program, Implemented on the Time Sharing IBM System 360, Model 50, in Teaching a Subject-Oriented User to Formulate Inquiry Statements to a Computerized On-Line Information Retrieval System
   by Dorothy Elaine Caruso

5. Applicability of the NASA Thesaurus to the File of Documents Indexed Prior to its Publication
   by Martha Manheimer

6. Pennsylvania Practicum Program for School Librarians: An Appraisal
   by Elizabeth A. Rupert

7. Information System Design Methodology Based on PERT/CPM Networking and Optimization Techniques
   by Anindya Bose

Publications available from

the University of Pittsburgh Book Center

Pittsburgh, Pennsylvania 15213

Acknowledgements

The author expresses gratitude to the chairman of his doctoral committee, Allen Kent, Director of the Communications Programs, and Director of the Knowledge Availability Systems Center, Professor of Library and Information Sciences, Professor of Computer Science, and Professor of Education, for his invaluable guidance and encouragement, and to members of the doctoral committee: Jack Belzer, Professor of Industrial Engineering and Library Science, and Associate Director, KAS Center; Frank B. Sessa, Chairman of the Ph.D. Program and Professor of Library Science, GSLIS; David I. Cleland, Associate Professor of Industrial Engineering; and Godfrey D. Stevens, Professor of Education.

While the author worked on the dissertation, his family patiently went through privations; no words of appreciation can adequately express his gratitude towards them.

Dr. Elaine Caruso intellectually shared some of the frustrations experienced by the author. She was a friend in need.

The cooperation of Mrs. Pearl Berger is gratefully acknowledged.

This work was supported in part by ONR Contract #N00014-67-A-0402-0004, Identifying #NR 049-258, and in part by the National Institutes of Health (Grant FR-00250).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. Statement of the Problem

"We have been faced over the last fifteen to twenty years with a technological revolution in the field of information handling which cannot really be compared to any previous technological revolution, at least in terms of the speed with which it has taken place. ... This has meant that the people who could control this technology have had to grow up with it over a very short period of time. Along the way, they had to develop the methodology to control this technological explosion."[1]

Concurrently, there has been an information explosion the like of which has never been experienced before and its exponential growth demands immediate development of methodology for its effective control.

There is a very close relationship between formalization in a discipline and application of information handling technology to it. In order to use the information handling technology, formalization must be introduced. Remarkable progress is being made in this direction as evidenced by the American Standard Code for Information Interchange (ASCII), the Federal Information Processing Standards (FIPS), the Medical Literature Analysis and Retrieval System (MEDLARS), the Machine Readable Catalog (MARC), now entering their second phases, efforts of the Library of Congress (LC), the National Library of Medicine (NLM), and the National Agricultural Library (NAL) for standardization and compatibility and enormous federal and foundation support in these efforts.

---

[1] Methodologies for System Design. Final Report on Contract no. AF 30 (602)-2620, Project No. 4594, Task No. 459-405. (Los Angeles: Hughes Dynamics, 1964), p. 1.5.

B. The Need for Methodologies

Today the need has become critical for well-defined methodologies to aid the information system designer and information system analyst in carrying out his tasks. This dissertation proposal arose from the recognition of the need for tools and well-defined methodology to aid in the process of information system design.

Over the past several years, we have seen the development of a technology which can aid in dealing with individual and isolated problems of information handling, e.g., file organization, remote access, digital transmission, etc. The information system designer faced with this wealth of potential physical tools has had no methodological tools on which to draw for analysis, evaluation, and synthesis of system based on the available solutions, except his own intuition and experiences. To my knowledge, not many works have specifically dealt with the problem of the tools and methods with which the information system designer does his work. There exists only the diverse experiences of individual workers. In part this is a result of the lack of recognition of system design as a major area of responsibility, and systems concept as a design tool.

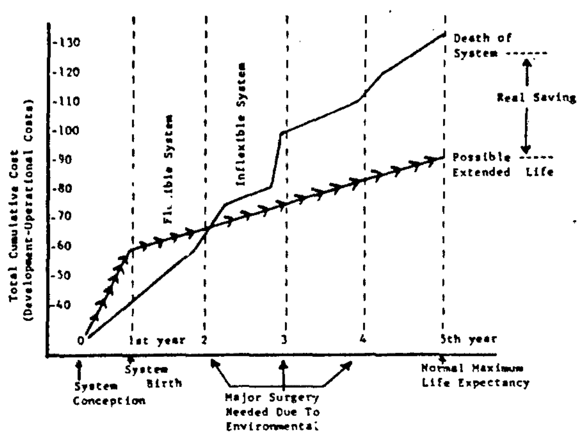
The problem bears upon deriving the methods by which the information system designer selects the particular solutions to the technical problems of system design itself.

C. What is Expected of the Methodology

The methodology should provide the techniques by which the information system designer identifies the system components and determines what techniques and equipment capabilities are required and then bring them into a functional and structural relationship. The guidelines and criteria for the selection of the equipment such as cost, equipment sophistication, manufacturer's reliability or other considerations, that is, the details of the processes by which the equipment may be specified or selected, should also be provided by the methodology. The design methodology should also bring into relief those areas where extra-design considerations must play a vital role, such as interfaces with other systems, physical or environmental limitations, etc., where decisions will have to be made by the policy makers.

D. Rationale for System Design

The design of an information handling system must be based on a detailed evaluation of the optimum combination of software, hardware and people, among other things, to guard against getting a system that is so constrained that it cannot grow when growth is required or a system that cannot change when change is necessary. Design of a system cannot always be judged on the basis of current work load or performance. Its survival potential against environmental, organizational and component changes is more important. The following figure 1 illustrates the



Flexible/Inflexible System

Figure 1

point that a flexible system (fish-bone line) which may require a slightly higher first year cost, might result in a distinct cost advantage over a hypothetical span of five years because the flexible system can accommodate changes more economically than the inflexible one (solid line).[2]

E. The Proposal

The dissertation proposal is to develop a methodology of information system design that will help a system design team do the job the management would demand of it. Categorically stated they are:

(1) Analysis - to find out what is to be done

(2) Design - to find out how it should be done

(3) Programming - to make the system a reality
- implementation
- operations
- evaluation
- control

in other words, to provide a methodology for Resource Allocation, Time-Scheduling, Optimizing System Performance, System Evaluation, and Control of its Performance.

F. Development of the Hypothesis

Making a decision is a process of rational selection among possible alternatives. During the past several years operations research and other business management and control techniques have reduced uncertainty and guesswork from business decision making. Manufacturing and

---

[2] Tom Scharf, "Management and the New Software," Datamation, XIV, No. 4 (April, 1968), 52, 57, 59.

production managers have benefited greatly from the advances in methods of time studies, manufacturing simulation,and the entire area of automation. But to the people concerned with development and design, these tools were not of much help. However, in the last few years, a new tool, the Project Network Model, has been introduced which is especially useful in development and design. It forces planners and designers to confront and to solve problems and difficulties 'en before the start of the system. The two major variations of this el are Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM). These have been successfully applied to complex engineering projects.

I believe these techniques with necessary additions, alterations and revisions, can be developed into a methodology for information system design. This methodology will put information system design on a rational basis by allowing the designer to show the precedence and time cost relationships of the activities and events of a system network.

G. The Hypothesis

PERT/CPM methodology or some modified version thereof can be developed into an Information System Design Methodology.

H. Methodology

MEDLARS, a large-scale, computer-based operational information system which is now entering into its second generation has been selected f.r this study. This will assure the availability of some data on working experience and evaluation. For our purpose MEDLARS is an ideal candidate for selection.

The historical and logical antecedents of MEDLARS have been reviewed to put the system in perspective. Then the system has been structurally and functionally analyzed tracing the actual activity-events-precedence relationship and delivery restraints, eventually coming up with a network representation of the system. PERT/CPM Network Model, Assignment Model and Sequencing Model have been developed with an abstract relationship to the system. Computer programs for PERT/CPM have been written for the IBM 360/50. The literature of the design and control methods in general and PERT/CPM in particular has been briefly reviewed, and the areas relevant to this work have been indicated.

I. Limitations

It did not seem feasible nor necessary to study the entire MEDLARS system for this dissertation. Only the Subject Indexing component of the Input Subsystem has been presented here.

II. Review of PERT/CPM

A. What is PERT/CPM?

Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) are time estimation and cost optimization techniques, respectively. They have been interfaced together to create a planning, designing, scheduling, and controlling technique for R & D and construction projects. It is based on a networking technique which establishes the time, cost, and precedence relationships among the activities and events of the network.

B. Background and History

Morgan R. Walker of the construction division of the E. I. DuPont DeNemours Company and J. E. Kelley, Jr. of Remington Rand's UNIVAC section are credited with developing the Critical Path Method (CPM), in 1957. In that year this new method was employed by DuPont in the construction of a ten million dollar chemical plant. Reportedly, DuPont credits this new method with savings of $1 million dollars on maintenance projects at Louisville.[1]

Concurrently, in 1957, a research team was established by the U.S. Navy Special Project Office to develop a program evaluation technique for the Fleet Ballistic Missile Weapons System development effort. The research team was composed of representatives from the Special Projects Office, the management consulting firm of Booz, Allen and Hamilton, and the Lockheed Missiles and Space Company. Through the efforts of this team, the Program Evaluation and Review Technique (PERT) was developed and

---

[1]Frederick J. Zalokar, The Critical Path Method; A Presentation and Evaluation, (Schenectady, N.Y.: General Electric, May 18, 1964), p. 1.

implemented as a research and development project management tool for the Navy's Polaris Program.[2]

In managing the Polaris missile project, the Navy became concerned with techniques for evaluating its progress. A schedule had been established for its development, and a system was set up for reporting the status, progress, and problem areas in terms of accomplishment or slippage (actual or predicted) of important program milestones. Major components were also evaluated and their status indicated by one of the following terms: "in good shape," "minor weakness," "major weakness," or "critical." These evaluations provided no measure of the impact on the overall program made by accomplishing a milestone or changing the forecast for its accomplishment. Tight schedules had been established for the program, so it was necessary to know the significance of a slip in a scheduled date, its impact on future scheduled dates, and the prospect for future slippages so that corrective action could be taken. As the slips in schedules and the prospects for future slips were studied, ". . . it appeared that the capacity to predict future progress was more limited than desired."[3]

As mentioned before, the operations research team was formed of representatives from the Naval Special Projects Office; Booz, Allen, and Hamilton, Inc.; and Lockheed Missile Systems division. This team was to study the application of statistical and mathematical methods to planning,

---

[2]Bruce N. Baker and Rene L. Eris, An Introduction to PERT-CPM, (Homewood, Illinois: Richard D. Irwin, Inc., 1964), p. 1.

[3]D. G. Malcolm, et al., "Application of a Technique for Research and Development Program Evaluation," Operations Research VII, No. 5 (Sept.-Oct., 1959), 646-669.

evaluation, and control of the Polaris program. The following objectives were established:

(1) To develop a methodology for providing the integrated evaluation of progress to date and the progress outlook, changes in the validity of the established plans for accomplishing the program objectives, and effects of changes proposed for established plans;

(2) To establish procedures for applying the methodology, as designed and tested, to the overall FBM (Fleet Ballistic Missile) program.

The team felt that the two major requirements for a program evaluation methodology were (1) detailed, well-considered time estimates for future activities, and (2) precise knowledge of the required or planned sequence in which the activities were to be performed. Since the time required to perform development activities is often uncertain, a procedure for quantitatively expressing this uncertainty was desired; this led to the statistical estimation technique, which is a primary feature of PERT. The sequence requirement was fulfilled by use of network plans.

PERT, therefore, was originally developed as a technique for evaluating established plans and schedules, but its utility is not limited to this. PERT can also be used as a planning and scheduling technique. The PERT technique for estimating elapsed times provides a way of handling some of the uncertainties in estimating the time required to perform many types of activities.[4]

---

[4]David I. Cleland and William R. King, Systems Analysis and Project-Management (New York: McGraw-Hill Book Company, 1968), pp. 279-280.

C. The Acid Test

A project on which Real Estate and Construction Operation has successfully used CPM was General Electric's Progressland Exhibit at the 1964-1965 New York World's Fair. W. F. Reardon, Regional Construction Manager, who had responsibility for the design and construction of the building and show portions of the Fair Exhibit, pointed out that this project was the acid test for CPM. "We had an opening day--April 22, 1964--which had to be met." Actually Real Estate and Construction Operation had adopted a CPM schedule with a completion date of March 22, 1964, saving a month for debugging and last minute items. Mr. Reardon attributes a great deal of the credit to CPM for having Progressland ready to roll on March 26, 1964, only four days off the target date of March 22, 1964.

Speaking from experience, Mr. Reardon states that the theoretical benefits of CPM are real benefits. "The critical activities were brought into the foreground and we knew exactly in which areas to concentrate our efforts to keep on schedule." To control the Fair project, Mr. Reardon organized bi-weekly construction meetings attended by Real Estate and Construction Operation, Turner Construction Company (general contractor for the Fair project), Walt Disney's organization and other interested parties. Following each meeting, a CPM review was made where actual results and estimated changes were developed for computer input. By the next morning, a revised CPM schedule was available for management's review. Mr. Reardon stressed the point that within twenty-four hours he could see how the decisions made at the construction meeting affected the total project and indicated that he was definitely sold on CPM.[5]

---

[5] Zalokar, op. cit., pp. 27-28.

D. Project Planning and Control[6]

Network plans are developed by first studying the project to determine the approach, methods, and technology to be used and then breaking it down into elements for planning and scheduling purposes. The elements of a project can be classified as follows:

(1) Project objectives. These are the goals to be accomplished during the course of the project. In most cases, the project objectives are specified before the plan is prepared; the plan merely prescribes the course to be followed in achieving the objectives.

(2) Activities, tasks, jobs, or work phases. These elements identify and describe the work to be performed in accomplishing the project objectives. They normally utilize time and other resources.

(3) Events or milestones. These are points of significant accomplishment--the start or completion of tasks and jobs, the attainment of objectives, the completion of management reviews and approvals, etc. They are convenient points at which to report status or measure and evaluate progress.

After the elements of the project have been determined, they are arranged in the sequence preferred for their accomplishment. This is a synthesis process that must consider the technological aspects of the activities and tasks, their relationships to one another and to the

---

[6] This section is partially based on Cleland, op. cit., pp. 270-285.

objectives, and the environment in which they will be performed. A network is used to reflect these factors as it portrays the sequence in which the project elements will be accomplished.

Networks are composed of events which are represented by nodes interconnected by directed lines (lines with arrows) which represent activities. Constraints are also represented as directed lines. Elements of the network correspond to elements of the project as follows: points in the network represent project objectives, with the direction of the lines indicating a precedence or sequential relationship; and directed solid or dashed lines indicate constraints.

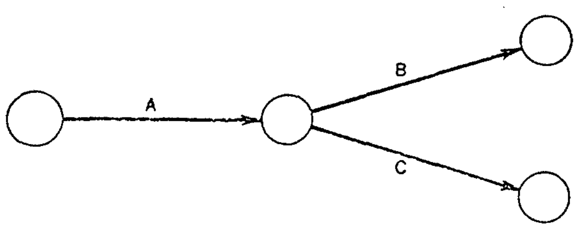
Activities are the jobs and tasks, including administrative tasks, that must be performed to accomplish the project objectives; activities require time and utilize resources. The length of the line representing an activity has no significance (in contrast to Gantt charts, where it is the significant factor). The direction of the line, however, indicates the flow of time in performing the activity.

Events are usually represented by small circles or squares. Numbers are used to identify the events and the activity that connects two events. Events represent particular points or instances in time, so they do not consume resources; the resources to accomplish an event are used by the activities leading up to it.

Constraints in network plans represent precedence relationships resulting from natural or physical restrictions, administrative policies and procedures, or management prerogatives, and they serve to identify activities and events uniquely. Constraints, like activities, are represented in a network plan by directed lines. However, constraints indicate precedence only; they do not require resources and normally do not require time. Those constraints which require neither time nor resources are represented by broken directed lines which are often referred to as "dummy" activities.

E. Preparation of Network Plans

The network plan is constructed by drawing directed lines and circles in the sequence in which the activities and events are to be accomplished.[7] The network begins with an event called the origin, which usually represents the start of the project and from which lines are drawn to represent activities. These lines terminate with an arrow and a circle representing an event, which may be the completion of a project element or an activity. All activities that are to be performed next are then added to the network plan by drawing a directed line from the previous event. For example, suppose activities B and C are to be performed upon completion of activity A. These three activities and their precedence relationship would be represented in the network plan as indicated in Figure 2. Activities and events are then added until the



Simple network plan.

Figure 2.

---

[7]There are two general methods which are used in actual construction of a network plan. This section describes the forward method, where construction begins with the start event and activities and events are added in sequential fashion to reach the end event. In the backward method, construction begins with the end event and proceeds backward to the start event. The backward method of network construction is often preferred to the forward method because attention is directed to the project objectives. With the objectives firmly in mind, the activities and events required to accomplish those objectives are often more easily determined.

project is complete. Constraints are added where required. The network plan terminates with one or more events, called terminal events.

To progress from one event to the next requires that an activity be performed. Each activity begins and ends with an event. The event at the start of an activity is called a predecessor event, and that at the conclusion a successor event. Time flows from a predecessor event to a successor event, as indicated by the arrow, and is normally from left to right throughout the network. As each activity is added to the network, its relationship to other activities is determined by answers to the following questions:

(1) What activities must be completed before this activity can start?

Activities that must be completed first are predecessor activities.

(2) What activities can start after this activity is completed?

Activities that can start after are successor activities.

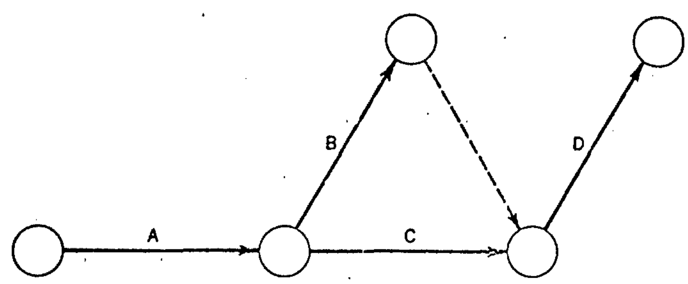
(3) What activities can be performed at the same time as this activity?

Those activities are concurrent, or parallel, activities.

In preparing the network plan, administrative activities must be included, such as the preparation of contracts, the procurement of parts, and the preparation of test procedures, specifications, and drawings. Technical work often cannot begin until a contract has been awarded or long-lead-time articles have been procured. A test cannot be started until specifications and drawings have been prepared and approved.

Two activities with a predecessor-successor relationship are called <u>sequential</u> activities. Performing activities in sequence requires that the start of the successor activity depends upon completion of the predecessor activity. Activities performed concurrently must be independent of one another. Independent activities may have a common predecessor event or a common successor event, but not both.

Suppose, for example, that activities B and C can be performed concurrently but that both are dependent upon the completion of activity A; activity D can be started after both B and C are completed. The relationships would then be represented as illustrated by Figure 3. The constraint, or dummy activity, is needed between activities B and D so



Network plan: correct predecessor-successor relationship
Figure 3

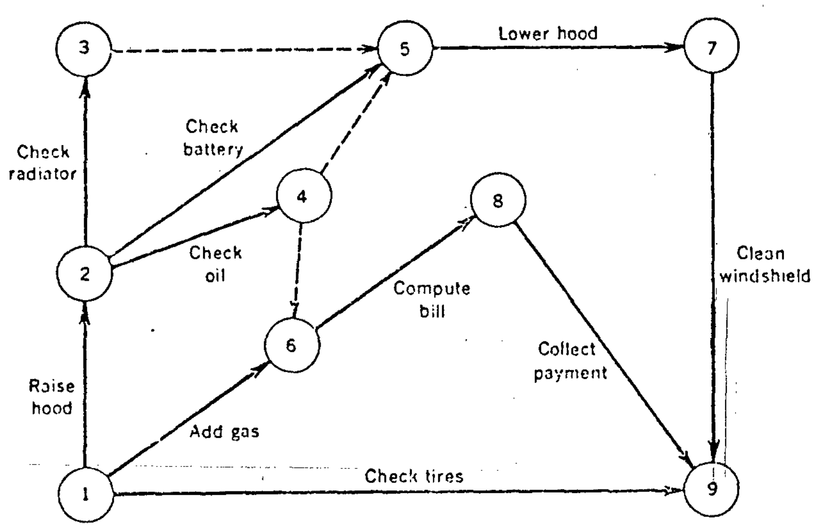
as to identify activities B and C uniquely by their predecessor and successor events.

F. An Illustration of a Network Plan

To illustrate the preparation of a network plan, let us consider as a project the servicing of an automobile at a service station. This example will be slightly exaggerated in order to emphasize the interrelationships between project activities that must be considered. The project situation is described as follows:

Automobiles arrive at a service station for gasoline. Services provided by the station include cleaning the windshield and checking the tires, battery, oil, and radiator. Sufficient personnel are available to perform all services simultaneously. The windshield cannot be cleaned while the hood is raised. Customers are charged only for gasoline and oil. Figure 4 shows the network plan. Events 1 and 9 are the origin and terminal events, respectively, representing the start and completion of service. Three constraints, or dummy activities, are used to sequence the activities properly.

The constraint between events 3 and 5, denoted as activity 3-5, is used so that the activities "check radiator" and "check battery" will not have common predecessor and successor events. The dummy activity 4-5 is used for the same reason. The constraint 4-6 is used to indicate that the activity of computing the bill cannot start until the activities "check oil" and "add gas" have been completed.



Network plan for servicing an automobile.

Figure 4

G. Analysis of Network Plans

The project network plan displays the activities, events, and constraints, together with their interrelationships. For the network to be useful in planning and controlling the project, time estimates must be made for the various activities which constitute the project.

A network path is a sequence of activities and events traced out by starting with the origin event and proceeding to its successor event, then to another successor event, etc., until the terminal event is reached. The length of a network path is the sum of the time estimates for all those activities on the path.

After activity time estimates have been made, an earliest and latest time for each event may be calculated. The earliest time for an event is the length of the longest path from the origin to the event. Thus, it indeed represents the earliest time at which the event can occur (relative to the timing of the origin event). The earliest time for the terminal event is the length of the longest network path. It therefore represents the shortest time required to complete the entire project.

The latest time for an event is the latest time at which the event can occur relative to the timing of the terminal event. If one imagines that the direction of each activity is reversed, the latest time for an event is determined by the length of the longest path from the terminal event to the event in question.

In calculating earliest event times, the general practice is to consider that the origin event occurs at time zero. The earliest time for each event is the sum of the earliest time for the predecessor event and the time for the predecessor activity. If an event has more than one predecessor event, this calculation is made for each of them, and the largest sum is selected as the earliest time for the event. This is so because the earliest time is the length of the longest path from the origin to the event.

To calculate the latest time for an event, the latest time for the terminal event is usually initially set equal to the previously computed earliest time for the terminal event. Then, for each event, the time for its successor activity is subtracted from the latest time for its successor event. The result is the latest time for that event. If an event has more than one successor event, this calculation is made for each, and the smaller result is used as the latest time for the event. This is compatible with the view of the latest time for an event as the longest path from the terminal event backward to the event in question.

Using these basic activity, event, and path measures, a number of network measures may be developed to aid in network analysis.

Event slack is the difference between the latest time and the earliest time for an event. The slack for an event is the difference between the length of the longest network path and the length of the longest network path through the event. Hence, event slack is a property of a particular network path.

The most important use of event slack is in identifying the critical path. The critical path is the longest network path. Thus, its length determines the minimum time required for completion of the entire project. Critical events are those events on the critical path. To identify critical events, one need only determine those events with the smallest amounts of event slack. Their identification is usually sufficient to identify the critical path; however, it need not uniquely identify it.[8] The operational significance of the critical events is that

---

[8] See Thomas L. Healy, Project Administration Techniques (Dayton, Ohio: The National Cash Register Co., April 1, 1963), for details of those special situations in which this may be the case.

they are the pacing elements of the project. If the project is to be expedited, the accomplishment of at least one of the critical events must be expedited. If there is a delay in the actual accomplishment of <u>any</u> critical event, the completion of the project will be delayed.

H. Using Network Plans in Planning and Controlling a Project

The construction of a network plan is a part of the planning function of project management. Network analysis makes use of the project plan to aid in scheduling a project.

Whether one is planning, scheduling, or controlling a project, the central idea involved in using network plans is the principle of <u>management by exception</u>. Stated simply, this means that it is the exceptions which require the attention of management. In the case of a project, the exceptions are the activities on the critical path, for it is they which pace the completion of the project.

If a project is to be expedited, some way must be found to hasten the accomplishment of critical events. Moreover, if the project is under way and the events on the critical path are not being accomplished according to plan, the project will be delayed if no way is found to hasten the completion of other critical events.

The application of the principle of management by exception in such projects usually takes the form of reallocating resources from noncritical activities to critical ones. This may be accomplished in either the planning or the control phase of the project; i.e., it may be done so that an earlier project completion date can be set up, or it may be done because the project is falling behind schedule. Presumably, such reallocations will permit faster accomplishment of critical activities and hence, faster completion of the project itself.

A number of techniques have been developed for accomplishing these ends. Among them CPM (Critical Path Method), PERT-Time and PERT-Cost are the best known and most widely used. After the network is prepared, the PERT planners obtain three elapsed time estimates for each activity: the shortest, the longest, and the most probable. These three estimates are used to compute the expected times required to perform each activity and a measure of the probability of accomplishing the activity in that time. The expected time estimate for each activity is used in analyzing the network. Variabilities in activity times are accumulated along the network paths in the same manner as activity times are accumulated, and they provide a measure of variability for each event. The variability associated with an event can be used to make statistical inferences about the occurrence of the event at a particular time, such as: the likelihood that the project will be completed by its scheduled completion date is 34 percent.

The PERT approach requires obtaining the activity time estimates from the people who are responsible for performing or for supervising the performance of the activities. The person directly responsible for the activity should be asked to make the estimate because he is most knowledgeable concerning its inherent difficulties and the variability in its accomplishment. Scheduled times cannot be used because they are not adequately responsive to changing conditions, contain no information on variability, and are often made under conditions and in an environment that do not reflect the technical aspects of the activity. A single elapsed time estimate would not, by itself, provide a measure of the variability in the time; this requires a range of estimated elapsed times. Estimates of the extreme times, reflecting the optimistic and pessimistic times, can usually be given with some degree of reliability, however, and it is felt that the most likely time estimate lies somewhere within this range.

The three elapsed time estimates, referred to as the optimistic, the most likely, and the pessimistic times, are defined below:

OPTIMISTIC TIME is the shortest time in which the activity can be accomplished. There should be practically no hope of completing the activity in less time than this, but if everything goes exceptionally well, it should be possible to accomplish it in approximately this time.

MOST LIKELY TIME is the normal or most realistic time required to accomplish the activity. If the activity were to be repeated numerous times under the same conditions and without any "learning-curve" effects, it would be accomplished most frequently in this time. (The most likely time is not the expected time, but an estimate based on experienced judgment; the expected time is a mathematically computed value.)

PESSIMISTIC TIME is the longest time required to accomplish the activity assuming unusually bad luck (e.g., major redesign or major reshuffling of planned action). The pessimistic time estimate should include such possibilities as initial failure and a second start, but not major catastrophic events such as strikes, fires, tornadoes, etc.

The range between the optimistic and the pessimistic time estimates is used in PERT as a measure of the variability of uncertainty in accomplishing an activity. If there is no uncertainty, all the time estimates will be the same, and the range will be zero. If there is considerable uncertainty, the range will be large. The time estimates must necessarily be based on planned assumed resources. The most likely time estimate must be based on the same level of resources that is used for estimating the optimistic and pessimistic times. For example, the optimistic time estimate must not be based on an extra shift or additional personnel, while the most likely time estimate is based on a normal shift and fewer personnel.

The most likely time estimate should be made first so that the estimate considers the available or planned level of resources and appraises the technical aspects of the activity realistically. The optimistic estimate can then be made, based on the same resources but with the assumption that everything goes exceedingly well. The pessimistic time estimate is made last, assuming that problems arise. The time estimates for each activity must be made independently and should not include a pad to cover possible delays.

An important property of the computed expected times is that they are added to calculate an earliest time, and this earliest event time is also an expected event time and has a probability of 50 per cent. This probability would not hold if most likely time estimates were summed in a similar fashion.

I. Efficacy of PERT

PERT has attracted considerable attention, which, to date, has probably been more extensive than its range of applications. The following comments and criticisms provide a measure of understanding of the basic technique.

Many feel that because the three time estimates are subjective, the estimator's personal bias will be introduced.[9] A fundamental principle of PERT is that the three estimates are to be made by persons who are most familiar with the technical aspects of the activities and therefore are best qualified to make the time estimates reflecting uncertainties involved in technical activities. Asking for three time estimates tends to remove the psychological barrier often encountered when only a single estimate is given, since a time range does not imply a commitment such as a single estimate does, and allowing the estimator to make a pessimistic time estimate permits him to provide for unforeseen contingencies that would probably be included as a pad in a single estimate. The effects of personal biases are felt to be cancelled in the analysis of the network, since estimates of optimists are offset by estimates of pessimists.

Another controversial aspect of PERT pertains to use of computed expected times for scheduling. It can be shown that PERT assumptions provide optimistic expected times. Therefore, many feel that scheduled times should be later than computed expected times. But some argue that automatically setting schedules later than expected times may increase the likelihood of schedule slippages and that expected times should not be automatically used for establishing schedules. The basis for this argument is that the computed expected times provide for slippage, and since roughly half the activities will be completed in less

[9]W. R. King and T. A. Wilson have hypothesized that a historical analysis of time estimating behavior can lead to the development of adjustment models. Such models could be used to adjust time estimates on the basis of historical estimating behavior. The adjusted estimates would presumably be superior to unadjusted ones. See "Subjective Time Estimates in Critical Path Planning: A Preliminary Analysis," Management Science, XIII, No. 5(January, 1967).

than their expected times and half will require more than their expected times, one will balance out the other. In actuality, however, R&D activities usually take as long as their schedules permit and are seldom completed ahead of schedule. Thus, schedule slippages occur in R&D activities which were not contemplated when schedules were prepared.

The validity of PERT expected time is another controversial matter. Where PERT is applied to the early stages of weapons-system development programs, the critical path is frequently 1 1/3 to 2 times as long as the originally planned program. No doubt the greater attention to detail that is necessary in applying PERT accounts for part of the additional time. A study of completed Air Force weapons-system development programs conducted independently of any PERT considerations, however, indicated that extensions of development time by one-third to one-half over the originally planned program were the rule rather than the exception.[10]

J. Advantages of Networking

Predicated on practical experience in the use of critical path methods, the following advantages have been observed:[11]

(1) provides a stimulus for long-range planning with considerable detail;

(2) facilitates the documentation and communication of the planning and control elements of a complex project;

---

[10] A. W. Marshall and W. H. Meckling, Predictability of the Costs, Time and Success of Development, Paper P-1821 (Santa Monica, Calif.: RAND Corp., Dec. 11, 1956).

[11] A. G. Holzman, "Critical Path Methods," Encyclopedia of Library and Information Science, ed. by Allen Kent and Harold Lancour, V(New York: Marcel Dekker, Inc., in press).

(3) projects the critical path through the network, thus permitting management to concentrate on the 10 to 20 per cent of the total activities which require the most judicious evaluation of the resources (management by exception);

(4) determines the impact on the total system resulting from a change in the original allocation of time and/or money.

K. PERT/CPM and Other Management Tools

There are quite a number of management tools that are available to a manager for project planning and scheduling. These are all useful techniques but they all have drawbacks and inadequacies, particularly when we come to the handling of projects, plans, and designs involving large numbers of interdependent activities, mutually dispersed in time and space, and having an element of uncertainty associated with most of them. PERT/CPM has been developed to handle this kind of problem.

The predecessor of PERT/CPM is the Gantt Chart, named after one of the early pioneers of scientific management, H. L. Gantt. The Gantt chart, also known as bar chart, is one of the most widely used planning techniques. It consists of a number of bars plotted against a calendar scale, each representing the beginning, duration, and end of some part of the total project. Though widely used, it has some serious drawbacks. These include:

(1) the lack of recognition of the interdependencies which exist between the efforts represented by the bars;

(2) the static scale which makes it difficult to reflect easily the dynamic nature of changing plans; and

(3) the inability to reflect uncertainty or tolerances in the estimation of time.[12]

But most of these difficulties can be solved by using PERT/CPM. The network approach of PERT/CPM makes it possible to indicate the interdependencies that exist between activities represented by the bars. Bar charts indicate which activities are currently behind schedule, but the downstream impact of these slippages on other activities cannot be readily ascertained, nor can the criticality of some activities be identified. The critical path approach of PERT/CPM enables the manager to concentrate his attention to the critical activities and reallocate resources if necessary. The statistical technique used to compute the "expected time" of an activity lets PERT/CPM handle the problem of uncertainty and identify the critical path through the network. Evolution of the bar chart technique to the network plan technique is illustrated in Figure 5.[13]

Figure 5A shows a number of bars plotted against a calendar scale, each representing the beginning, duration, and end of some part of the total project. The small arrowheads point to some milestone events. From this figure one would not get any idea as to how the bars interrelate to each other and how an interrelationship is going to affect the project as a whole and how optimally a slippage could be handled, should one occur. Figure 5B transforms the bars into activities (lines) and events (squares).

Figure 5C establishes interdependencies between the events at a relatively macro level, and Figure 5D takes it to a relatively micro level,

---

[12]Russel D. Archibald, PERT Management Information Systems (Culver City, Calif.: Hughes Aircraft Corporation, 1952), p. I-1.

[13]Adapted from Baker and Eris, op. cit., pp. 54-55.

EVOLUTION OF THE BAR CHART TO THE NETWORK PLAN CONCEPT

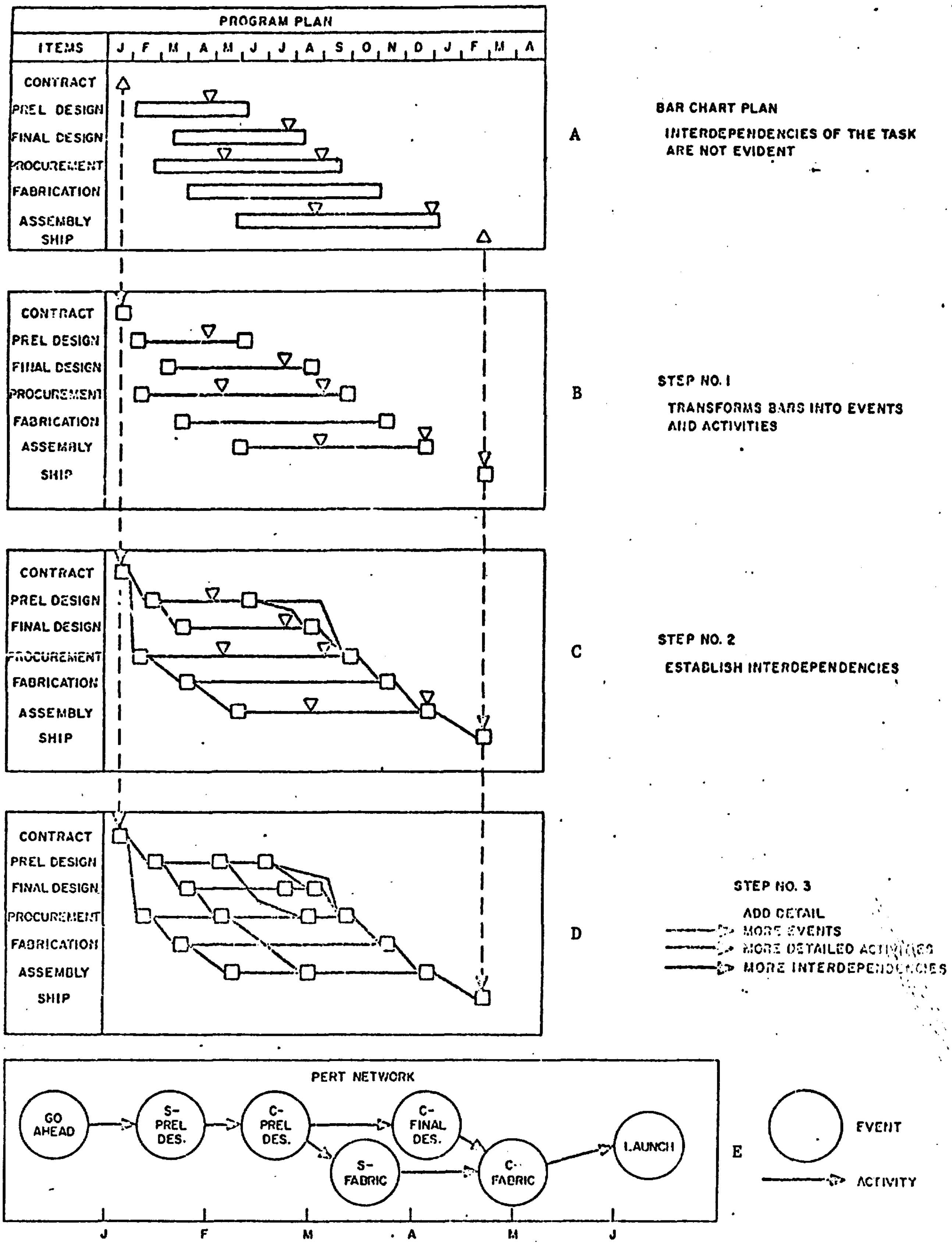


Figure 5.

adding more detail, incorporating more events and activities, and showing more interdependencies. Finally, Figure 5E shows a simple PERT network. The S's and C's inside the events mean 'start' and 'complete', respectively. With its time estimation and cost computation capabilities coupled with the ability to identify the critical activities and path through the network, PERT/CPM has become a very powerful technique for both R & D, as well as project planning, scheduling, and control; and above all, it lets one manage by exception. However, this need not preclude us from using other techniques in conjunction with PERT/CPM to complement one another, such as line-of-balance (LOB) discussed on page 30.

A typical family of networks is illustrated in Figure 6, showing a successive blow-up technique of activities between milestones.[14] Figure 6A shows a summary of major milestones of the project; these milestones are important target events of the project such as completion of a software package or assembly of a hardware system. This summary network is analogous to our "Umbrella Net" (see page 120).

At the level 1, Figure 6B, some activities between milestones, rather than just milestone to milestone link, have been indicated. At the level 2, Figure 6C, the activity 1-4 of level 1 has been expanded (Chart 2-A); the same has been done for activity M4-7 and 7-9 of level 1 (Chart 2-B).

At the level 3, Figure 6D, the activities of level 2 have been expanded as follows: activity 1-5 and 8-4 of chart 2-A in chart 3-A and 3-B, respectively; and activity 2-7 and 5-6 of chart 2-B in chart 3-C and 3-D, respectively.

---

[14] Ibid., p. 46.

TYPICAL FAMILY OF NETWORKS

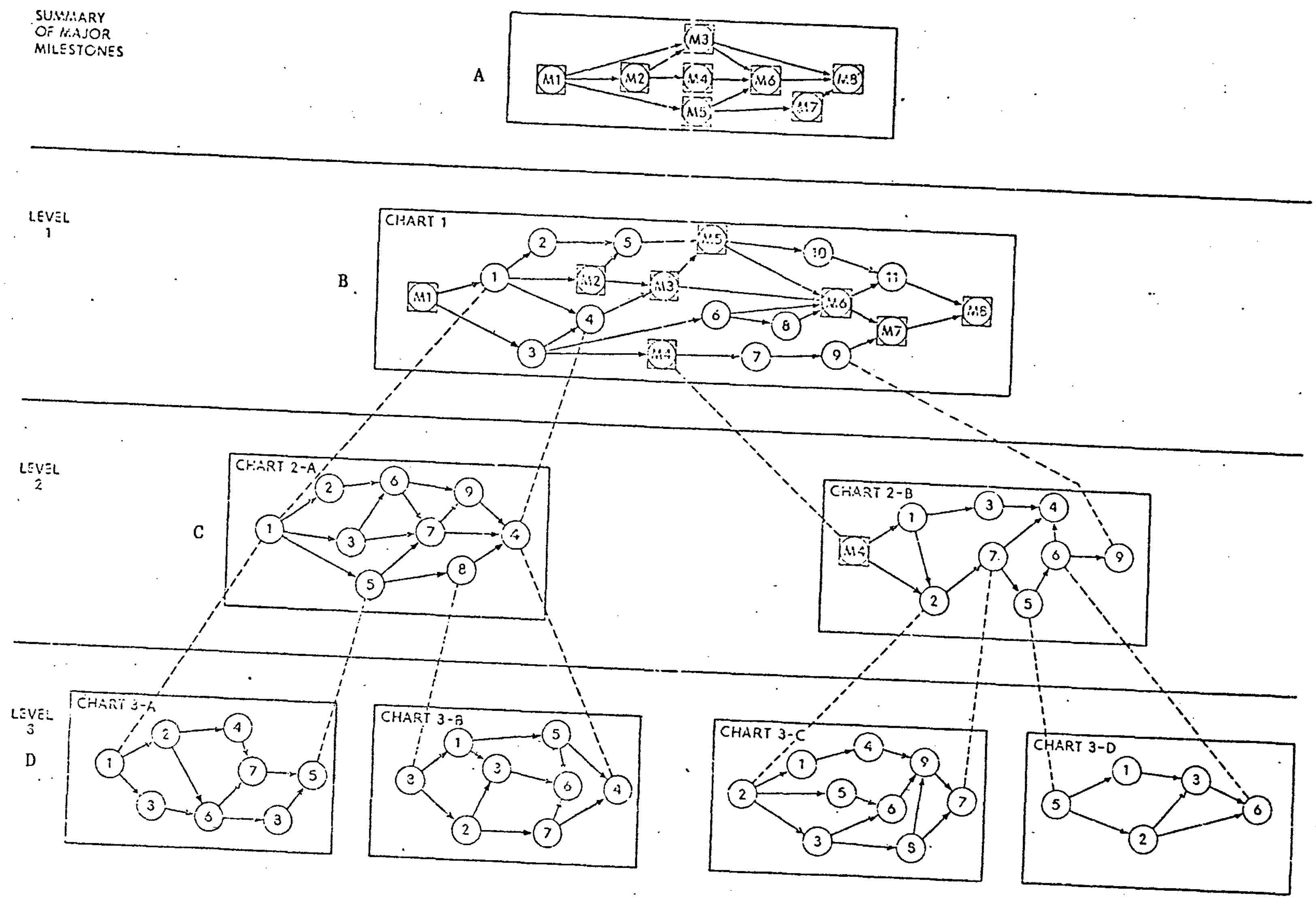


Figure 6.

Some of the different planning and scheduling techniques, besides PERT/CPM, that exist today, are illustrated in Figure 7.[15] Figure 7A is a Bar and Event chart plotted against a calendar scale, showing progress of the project (solid area).

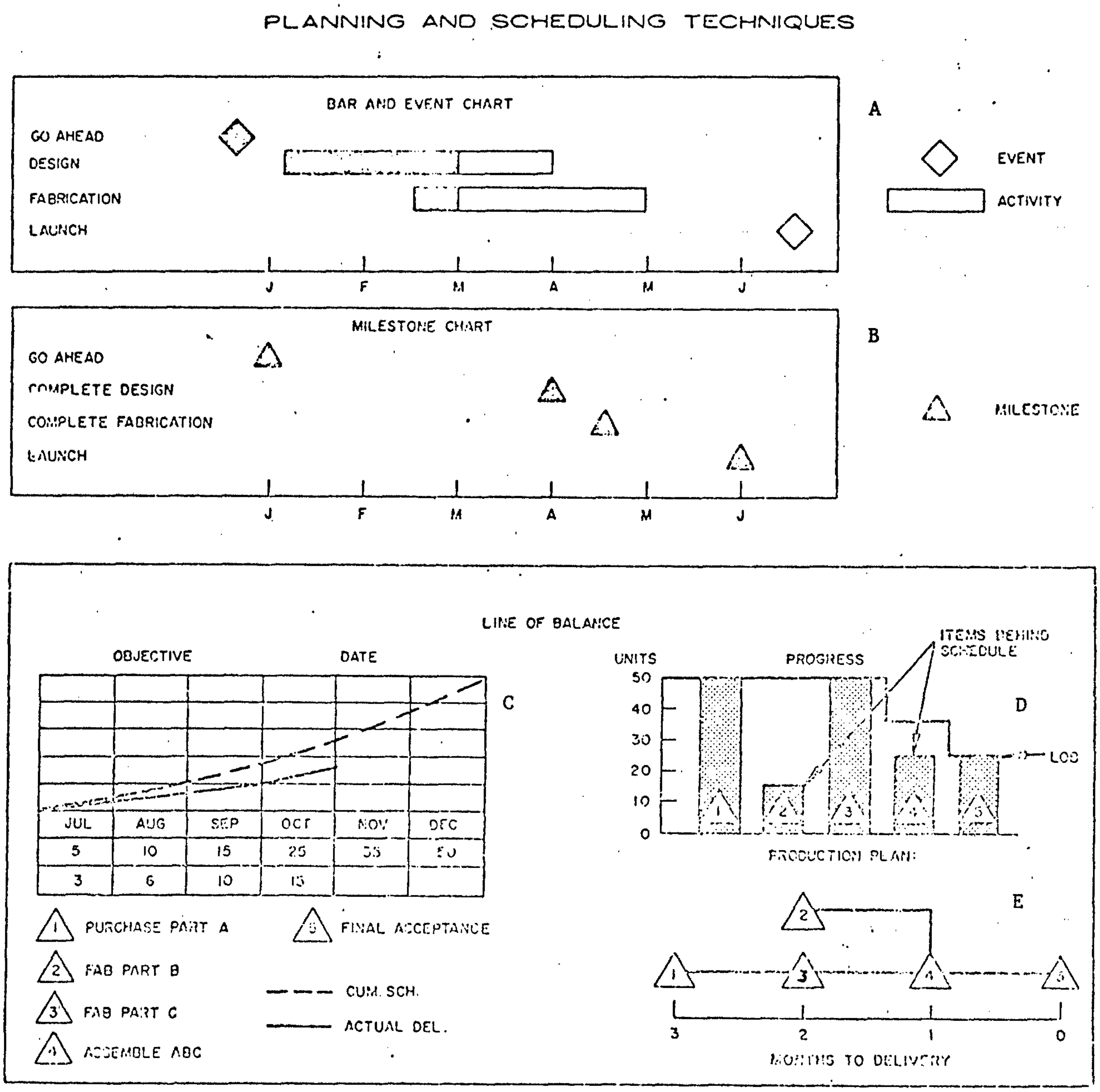
A milestone chart shows the significant project event or milestone in chronological order to form a diagonal from left to right on the chart, Figure 7B. This technique suffers from similar drawbacks as the bar chart. It lacks the ability to measure the impact of slips and changes on the total project or to adequately differentiate between critical and noncritical problem areas.

Line of Balance (LOB)[16] is a production planning and control system which time-schedules key events necessary for completing an assembly (Figure 7D), with respect to the delivery dates for the completed system. This management tool uses graphic displays to monitor the progress of production contracts. Production plan progress is bar charted (Figure 7D, showing items 2 and 4 behind schedule, and the LOB), and compared with the production objective which is in graphic form (Figure 7C, showing cumulative schedule, broken line, and objective numbers, 2nd row from the bottom; actual delivery, solid line, and numbers representing actual delivery, bottom row) and a line of balance is generated to show revised requirements for meeting the scheduled production plans. Figure 7E shows months remaining for delivery and uses this as a scale to show the flow and interrrelationship of the project events 1 through 5. The "management-by-exception" approach is used here to expose weaknesses in the production program so that correct action may be taken to eliminate the weak areas.

---

[15] Adapted from Baker and Eris, op. cit., p. 55.

[16] Ibid., p. 56.



PLANNING AND SCHEDULING TECHNIQUES
BAR AND EVENT CHART
GO AHEAD
DESIGN
FABRICATION
LAUNCH
J F M A M J
A
EVENT
ACTIVITY
MILESTONE CHART
GO AHEAD
COMPLETE DESIGN
COMPLETE FABRICATION
LAUNCH
J F M A M J
B
MILESTONE
LINE OF BALANCE
OBJECTIVE
DATE
JUL AUG SEP OCT NOV DEC
5 10 15 25 35 50
3 6 10 15
C
UNITS
PROGRESS
ITEMS BEHIND SCHEDULE
50 40 30 20 10 0
D
LOB
PRODUCTION PLAN
E
3 2 1 0
MONTHS TO DELIVERY
PURCHASE PART A
FAB PART B
FAB PART C
ASSEMBLE ABC
FINAL ACCEPTANCE
CUM. SCH.
ACTUAL DEL.

Figure 7.

Initially the objectives of PERT and CPM were extremely divergent. CPM was developed within the construction industry where previous experience in similar work can be used to predict time duration and cost within a range. While many of the characteristics of PERT and CPM are the same, one of the essential differences is that PERT recognizes that the actual activity times are not deterministic, but instead, may have considerable chance variation. CPM, on the other hand, ignores the chance element associated with the activities and employs only normal and crash cost/duration for each activity.

As we have seen, PERT was originally designed to plan and control large systems implementation where little past experience has been accumulated. A typical example of PERT would be the research and development required to structure an information system to transfer NASA space technology to industry. No experience was available on information scientists, engineers, programmers, and computer hardware to implement such a system; therefore, it was probable that the times for activities in the network representing this system would have considerable variance. But in the construction of a new library, one could draw from the considerable experience of professional librarians and architects to obtain more reliable estimates of activity times.[18]

Since CPM has the capability of activity cost optimization, and PERT has the capability of activity time estimation, it seems logical that these two methods will be interfaced.[19] Thus, although these two methods were developed in different environments over the years, they can most profitably be used in conjunction for planning, design, scheduling and control.

---

[18] Holzman, op. cit.

[19] DoD and NASA Guide, "PERT Cost Systems Design" (Washington, D.C.: Office of the Secretary of Defense, NASA, June, 1962)

It appears to this worker, however, that to use PERT/CPM as a function control tool we need something more than just networks indicating precedence and time/cost relationships. A particular activity in a large and complex PERT/CPM network would not necessarily know what to do or where logically to go, if anything goes wrong. In a large and complex network an activity will be very small, and being preoccupied with its own activity, may not have the feel of coordinated belonging with the system as a whole.

When we have laid out a system in the form of a PERT/CPM network of interrelated components indicating the flow and precedence relationship of the system activities, we have a physical network, but it does not help us in understanding the logical or control relationship between the activities.

An activity is physically and sequentially related to its predecessor and successor activities but its logical or control relationship may be entirely different, and this relationship need not have to be sequential, or in tandem; i.e., activity n need not have to be logically related to n-1 and n+1, where n-1 and n+1 represent the predecessor and successor activities of n, respectively. As a matter of fact, an activity may be logically related to any other activity or activities in the network, depending on the control that has been established for the network.

Two different types of control are needed. These may be called intracontrol and intercontrol. Intracontrol may be defined as those control problems that may be handled within an activity, e.g., the finish of a product. Intracontrol cannot be separately represented in a network because it is ingrained in an activity. For this reason, a PERT activity has been modifed and redefined in this dissertation.

Intercontrol, however, can be separately represented and logically interwoven with the network without interfering with the immediate time/cost computations of the physical network. Intercontrol may be represented by links to control nodes and/or activity nodes. A control link to an activity node will mean a link to the intracontrol of the activity concerned. A control link between two sequentially adjacent activities is always assumed and coincides with the activity arrow.

The possible use of control couplers between nodes is well worth investigating.

III. ALLOCATION OF RESOURCES

A system is a network of interacting components organized to achieve some goal. Every component does something towards the achievement of system objectives. To do this every component must receive some input either from another component belonging to the system or from its environment. This input will be processed by the component concerned and generate an output which will be an input to some other component belonging to the system or to the environment.

The system components will be using up resources in this process. Here we are concerned with the resources which are available to the system and we will assume, not too unrealistically, that resources are limited. Under the circumstances, the objective is to allocate the limited available resources to the components so as to either minimize the total cost or maximize the total return.

A. Linear and Dynamic Problems

The problem is twofold. We may try for immediate optimization, or we may try for ultimate optimization. If we try for immediate optimization we are assuming effectiveness as linear functions of allocations. But if we intentionally decide to sacrifice a little bit at time t-1 to be in a better position at time t, we are assuming a dynamic relationship between allocation and effectiveness. Programming for optimum allocation may take a stochastic turn when current decisions are based on estimates of probable future values of parameters.

Most allocation problems can be represented by a matrix such as is shown in the following Table 1 .[1] The entries in the cells c_{ij} represent

[1] Russell L. Ackoff and Maurice W. Sasieni, Fundamentals of Operations Research (New York: John Wiley & Sons, Inc., 1968), p. 121.

Resources	Jobs to Be Done J_1	J_2	$\cdots$	J_j	$\cdots$	J_n	Amount of Resources Available
R_1	c_{11}	c_{12}	$\cdots$	c_{1j}	$\cdots$	c_{1n}	b_1
R_2	c_{21}	c_{22}	$\cdots$	c_{2j}	$\cdots$	c_{2n}	b_2
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
R_i	c_{i1}	c_{i2}	$\cdots$	c_{ij}	$\cdots$	c_{in}	b_i
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
R_m	c_{m1}	c_{m2}	$\cdots$	c_{mj}	$\cdots$	c_{mn}	b_m
Amount of resources required	a_1	a_2	$\cdots$	a_j	$\cdots$	a_n	

TYPICAL ALLOCATION PROBLEM

Table 1

the cost or return that results from allocating one unit of resource R_i to job J_j. The principal techniques available for solving allocation problems involve the assumption that the amounts of resources available (b_i), the amounts required (a_j), and the costs (c_{ij}) are known without error. As we know, this is not always the case. Hence it is sometimes desirable to determine how sensitive a solution to an allocation problem is to possible errors in these coefficients.

B. Optimization of Cut-back

If the sum of the available resources, $\sum_{i=1}^{m} b_i$ is equal to the sum of the resources required, $\sum_{j=1}^{n} a_j$, we have a balanced allocation problem. However, if

$$\sum_{j=1}^{n} a_j \neq \sum_{i=1}^{m} b_i,$$

we have an unbalanced problem that requires not only allocation of resources to jobs, but also the determination of either what jobs should not be done (if $\sum_{i=1}^{m} b_i < \sum_{j=1}^{n} a_j$), or what resources should not be used (if $\sum_{i=1}^{m} b_i > \sum_{j=1}^{n} a_j$); in other words, the optimization of a cut-back problem.[2]

C. The Problem is Generic

The complex of Assignment, Sequencing, and Distribution, and Optimum allocation of limited resources, constitute the set of generic problems which applies to most systems, including information handling systems.

In an Assignment Problem each job requires one and only one resource, and each resource can be used on one and only one job. This is a case where resources are not divisible among jobs, nor are jobs divisible among resources. An example of an assignment problem may be assigning men to offices or jobs, drivers to trucks, classes to rooms, or problems to research teams. The problem here is to find a unique one-to-one pairing of resources and jobs so as to optimize the performance of each pairing that is made. Where there are more jobs to do than can be done, it is possible to decide by applying the assignment technique which job to leave undone or what resources to add, to minimize cost or maximize return.

In a situation where resources can be divided among jobs, it becomes possible to do some jobs with a combination of resources. A problem that involves the distribution of empty freight cars to locations requiring

[2] Ibid., pp. 122-123.

them, or the assignment of orders to be filled to stocks at warehouses or factories, is a transportation or distribution problem. It is a problem of allocating resources from one or more sources to jobs needing them (destinations), when the jobs may be performed by combining resources from several points. Transportation or distribution technique makes it possible to add or subtract resources or jobs on a rational and quantitative basis.

Sequencing is the selection of an appropriate order in which to serve waiting customers or do jobs. A sequencing problem includes projects or jobs that consist of tasks that must be performed in a specified sequence. An example can be given in a "job shop" context. In a job shop, a production facility that processes many different products over a variety of combinations of machines faces a sequencing problem or, in other words, a scheduling problem. PERT/CPM is a networking and scheduling technique endowed with the capability of identifying the critical tasks that control the time required to complete the project and optimizing the time/cost relationship of the project activities. PERT concerns itself with the uncertainties in activity times under special conditions, like optimistic or pessimistic, but does not address itself to direct control of activity times by allocation of resources to tasks. It is the function of CPM to do this in a deterministic context.

At one time or another, an information handling system component will have to distribute or transport its equipment or facilities, assign jobs to capabilities, and sequence them in some order to optimize component and, ultimately, system performance.

The networking technique that PERT/CPM provides can, with some modifications, hold and represent continuously and in parallel, in a graphic form, the physical and actual "activities" (and "events") of the design and operation of an information handling system. In this case the activities are substitutes of system components reduced to the basic functional unit level.

IV. CHARACTERISTICS OF INFORMATION SYSTEMS

A. Definition of Information Systems

An information system is a set of interrelated components to meet a defined information need. It is essential to differentiate between an information system as such and the particular technology which, in a given time and place, is utilized as one feature of the system. There is a tendency, however, to classify types of systems by technological characteristics rather than by the characteristics of information systems.[1] Information systems should be designed around the informational needs of the system users rather than around available technology. The foundation of information system development is the analysis of the need for information at all levels and for all functions of the system users. This analysis of the user need must precede commitment to a particular type of equipment.

An information system should be capable of transferring information laterally across departmental lines as well as vertically through different levels of organizational hierarchy.

"When we look at the historical development and evolution of information systems . . . it becomes evident that classification schemes based on such criteria as 'scientific,' 'commercial,' 'real-time,' and 'off-line,' are too narrow for our purpose and too specific to particular technical design issues."[2] The typical information system encompasses some combination of these features. For our purpose, we shall define an

---

[1]Perry E. Rosove, Developing Computer-Based Information Systems (New York: John Wiley & Sons, Inc., 1967), p. 4.

[2]Ibid., p. 11.

information system as an integrated, multi-purpose, geographically localized or dispersed, computer-based configuration of people, procedures, data, and equipment, designed to satisfy the information needs of the system user.

B. Nature of Information Systems

An information system is tailor-made to fit the needs, objectives, and requirements of a user-group. Between information systems, among other things, there will be differences in computer programs, format and content of displays and reports, kinds and format of data base, relationships among system components, and in the mode of man-machine symbiosis.

Information systems are one-of-a-kind, that is, only one operational system is usually developed from the design. An information system is not a mass-produced article. A major consequence of mass production of an article is the fact that a complete prototype can be built before full-scale mass production, at a fraction of the total cost of the project. The prototype can be used to test and evaluate the design against specifications and performance criteria. If necessary as a result of the test and evaluation, the prototype can be modified without entailing considerable cost. Once the prototype model meets all user requirements, the design is frozen and production is started. But unfortunately, the creation of a complete prototype for an information system would be tantamount to producing the operational system itself and the cost of producing the prototype would be prohibitive and defeat the whole idea of producing a prototype.

Alternatives to prototype production for information system development are feasibility studies of system components and subsystems, and running a test facility under experimental and simulated conditions in which the basic design concepts of the new information system are tested. The creation of a test facility before the construction of the information system itself is illustrated by the Cape Cod System, which was built in 1953 as a working model of the SAGE system of air defense (Semi-Automatic Ground Environment).[3] At best, however, a test-bed information system can only represent a truncated version of the operational system. If the information system is of crucial importance such as in a defense system, there must be a backup system to take over in the event that the primary system is destroyed. Besides defense systems, air-traffic control systems, space surveillance systems, space-vehicle tracking and recovery systems, air-sea rescue systems, weather forecasting systems, fire warning and control systems, law enforcement systems, emergency ambulance systems and the like should have survivability, redundancy, alternate modes of operation, and backup capabilities.

Changes in information systems come as planned evolution. As the old system is phased out, the new system is phased in. The system as it exists at any stage or phase incorporates earlier phases. An information system is adaptive to its environment; it adapts itself to changing situations and learns from experience, thanks to its human components. Modifications to the system should be made through an on-going dialogue among the system designers, system operators, and users of the output of the system.

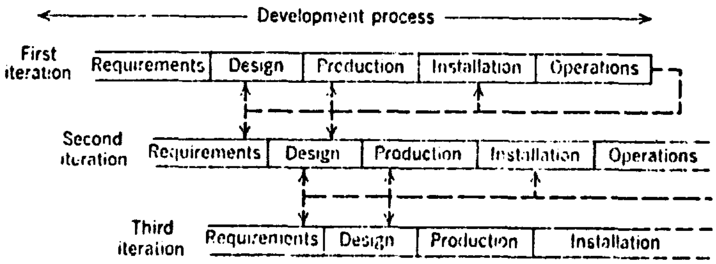
An information system design may not push the hardware, software, and human capabilities to the limit. The level of sophistication of an information system depends on managerial decisions rather than the state-of-the-art. Managerial desire to initiate with a modest capability, lack

---

[3] _Ibid._, p. 37.

of funding, and inadequate understanding of the user requirements may all be reflected in the design of an information system.

The evolutionary process of an information system is an iterative one. The first development cycle may be in the production phase, while the second is in the design phase and the third in the requirement phase (Figure 8).[4] This iterative, evolutionary character of development of information systems relies heavily upon the flow of data among the design personnel working at the different levels of iteration as shown by the dashed lines in Figure 8. The system cannot meaningfully evolve without the provision of a feedback system.



INFORMATION SYSTEMS PHASES

Figure 8

Although information systems could profit from improvements in such areas as core storage capacities, speed of operation, display devices, and input/output devices, the technological limitations in these fields do not constitute insuperable constraints on the design of contemporary information systems.

The computer is a basic component in large-scale information systems, but since humans also constitute other important components of the system and will keep doing so until content analysis, indexing,

[4] Ibid., p. 43

abstracting and the like can be thoroughly mechanized, and since humans will never be replaced as the ultimate recipient of the output of information systems, the designers of such systems have to get involved in the so-called "soft" sciences like human relations, management science, psychology, sociology and other behavioral sciences. These sciences and others, such as human engineering, are applied in the design of information systems to obtain an optimum symbiotic relationship between human and physical components of information systems.

It is difficult to determine the effectiveness that is bought for a dollar when the management is paying for an information system. More often the effectiveness is intangible; it is hard to assign a dollar value to it. We cannot live without air, but we do not pay every time we breathe; however, if we had to, we wouldn't know how to fix the price. Information services, likewise, are indispensible for civilized society but it is very difficult to put price tags on these services. Moreover, traditionally, information services have been offered more or less free of charges, and hardly ever have information services had to justify their existance by showing a profit or a favorable cost effectiveness ratio.

With the availability of customized "instant" information, thanks to the random access devices and time-sharing computer systems, and of machine processable, discipline or mission oriented data sets, the time has come to establish a value theory or price theory of information.

For the computer-based information systems, compatibility and interface with external and sometimes internal information systems are important concerns. The experience of MEDLARS with its national and foreign search centers, of the three federal libraries, and of the military command and control information systems, point to the necessity of handling the compatibility and interface problem as a design requirement.

C. Information System Development Process

The development or design of an information system is the creation of a new or a replacement system which is designed to meet the information needs of the system user. "System development is concerned with the entire history of a particular information system, including the study and analysis of its manual or semimanual predecessor; the initial conception of the replacement system; the analysis of existing user objectives and the creation, in consultation with the user, of new objectives; the definition of the new system's operational requirements; the design of the system; the specification of its physical components; and the production (or cause the production) of these physical components. Systems development includes provision for the human components of the system, that is, personnel and organizational design. It includes the creation of training programs and capabilities for system testing and system evaluation. And, given the concept of system evolution, systems development must also include over-all, long-range planning for the evolutionary replacement of each system configuration by subsequent ones."[5]

D. Systems Development Phases

In the course of its development, every large-scale information system must pass through a sequence of six stages in its life history, namely:

---

[5]Ibid., p. 17.

Phase I - Requirements

Phase II - Design

Phase III - Production

Phase IV - Installation (Implementation)

Phase V - Operation

Phase VI - Evaluation (Continuous)

E. Systems Engineering and Operations Research Approach

Goode and Machol describe the emergence of a systems orientation in the field of engineering. They point out that early efforts to develop large-scale equipment systems, such as the telephone system, applied methods and an approach which had worked well in the design of small-scale systems.[6] In the design of large-scale systems, this approach, however, was not successful since the components of the large-scale system did not work when they were joined together. Out of these early failures, there emerged new concepts and new methods, and the name "Systems Engineering" was given to the field. The method is the interdisciplinary team approach. The evolutionary forces which resulted in the development of systems engineering as a field in the 1950's were increasing system complexity and the growth of modern technology, which broaden the range of possibilities and alternatives. According to A. D. Hall, the systems point of view means that the systems engineer is not concerned primarily with the devices that make up a system, but with the concept of the system as a whole--its internal relations and its behavior in the given environment.[7] Systems

[6] H. H. Goode and R. E. Machol, Systems Engineering: An Introduction to the Design of Large-Scale Systems (New York: McGraw-Hill, 1957), pp. 7-8.

[7] A. D. Hall, A Methodology for Systems Engineering (Princeton, N.J.: Van Nostrand, 1962), VII, p. 16.

analysis and operations research are the distinguishing features of the field of systems engineering. Churchman asserts that operations research should equal "Systems Science."[8] Operations research is a technique developed in an effort to apply scientific method in systems problems. A central orientation to the solution of such problems is the systems approach, since the industrial organization is regarded as an interconnected complex of functionally related components.

Operations research and systems science had an impact on business, but relatively little direct influence on information systems development. Although the system concept existed as early as the 1940's, the development of integrated information systems in business appears to have been the result of a trial and error process.[9] At the present time systems science is gaining ground as a philosophic concept, and systems engineering as an operational tool. However, in the information science field the systems point of view has not yet prevailed in an operational, day-by-day sense.[10]

The problems that caused the development of systems science are the problems of precedence, dependence, and interrelations. There can be a situation where the components are working perfectly but the system as a whole is working at less than optimum efficiency. As an analogy, the different states of a nation may be in perfect harmony internally, but as far as their federal relationships are concerned, the nation may

---

[8]C. W. Churchman, Does Operations Research=Systems Science?, Symposium on Operations Research (Santa Monica, Calif.: System Development Corporation, March 27, 1963).

[9]Rosove, *op. cit.*, p. 13.

[10]*Ibid.*, p. 16.

be facing disruption. So it has become a primary concern for system designers to be able to "design" the relationships between the system components to assure system optimization and survival.

PERT/CPM is a networking technique. It lets the designer establish the precedence and dependency relationships between the system components. In a graphical representation of the system it brings into relief the "federal" relationships and allows the designer to do the necessary problem solving in the area of precedence, dependency, and interrelations.

Scheduling is the process of accepting input, operation on the input by assignment and/or sequencing, and producing an output. In a system this is the complement of precedence and dependency relationships, and the two together complete the picture. That is, for most systems, including information systems, scheduling is the activity side of the system and networking represents the interrelationship side of the system. So it seems logical that the networking technique of PERT/CPM and the scheduling techniques of assignment and sequencing should be interfaced to develop an information system design methodology that is capable of providing the designer with a gestalt approach so that he can design both the activities and the interrelationships of a system with properties not derivable from its parts in summation.

V. PROBLEMS OF INFORMATION SYSTEMS: GENERAL

We have established a definition of information systems and discussed their characteristics, cost/effectiveness, compatibility and interface problems, and their development processes and phases. But what are the problems that an information system would normally encounter in performing its design functions? Is it possible to develop design requirements from the diagnostics generated by the system operating experience and create design algorithms which will force the designer to go through the process of problem solving at the point of their logical occurrence on the drawing board? Is it possible to develop a design methodology which will also provide mechanisms for trouble shooting as they will occur at the basic functional unit level? Before we may attempt to answer these questions, we have to find out problems that are normally encountered by an information system. Then we will be in a position to consider the question of developing a design methodology that can live up to these problems.

According to Kent[1] any information retrieval system must carry out certain unit operations. These unit operations cover the whole gamut of information system activities starting from the identification and acquisition of information down to the delivery of search results. For the purpose of his book, Kent assumed the existence of the files of records and itemized the rest of the unit operations as follows:

(1) Analysis, involving perusal of the record and the selection of points of view (or analytics) that are considered to be

[1]Allen Kent, Textbook on Mechanized Information Retrieval (2nd ed.; New York: Interscience Publishers, 1966), pp. 20-22.

of sufficient probable importance to warrant the effort of rendering them searchable in the system.

(2) Vocabulary and subject heading control, involving establishment of some arbitrary relationships among analytics in the system. These arbitrary relationships are usually dependent on similarities among analytics as revealed in dictionary definitions for the words used to express the analytics.

(3) Recording of results of analysis on a searchable medium, involving the use of a card, tape, film, or other medium, on which the analytics are transcribed.

(4) Storage of records, or source documents, involving the physical placement of the record in some location, either in its original form, or transcribed or copied (in full or reduced size) onto a new medium.

(5) Question analysis and development of search strategy, involving the expression of a question or a problem, the selection of analytics based on analysis of the question, the expression of these analytics in terms of a particular search mechanism, and their arrangement into a configuration that represents a probable link between the question as expressed and the records on file as analyzed.

(6) Conducting of search, involving the manipulation or operation of the search mechanism in order to identify records from the file.

(7) Delivery of results of search, involving the physical removal or copying of a record from file in order to provide it in response to a request.

In the following flow chart, Figure 9, Lancaster[2] has summarized all the activities involved in the storage and retrieval process from the time a document is indexed for input to the system until it is retrieved and delivered to a user in response to a request made to the system.

Kent[3] identified the following procedure for the development and study of information systems:

(1) Identify the records, or source documents, that are to be (or have been) included.

(2) Decide on the extent, or depth, of analysis of the records that will match the probable extent, or depth, of questions that are to be put to the system.

(3) Select a system of terminology or subject heading control or coding that will match in precision that of the probable search.

(4) Select a suitable searching device or technique that will probably be useful and economical, and

   (a) select a system of notation for recording the results of analysis on the search medium; or

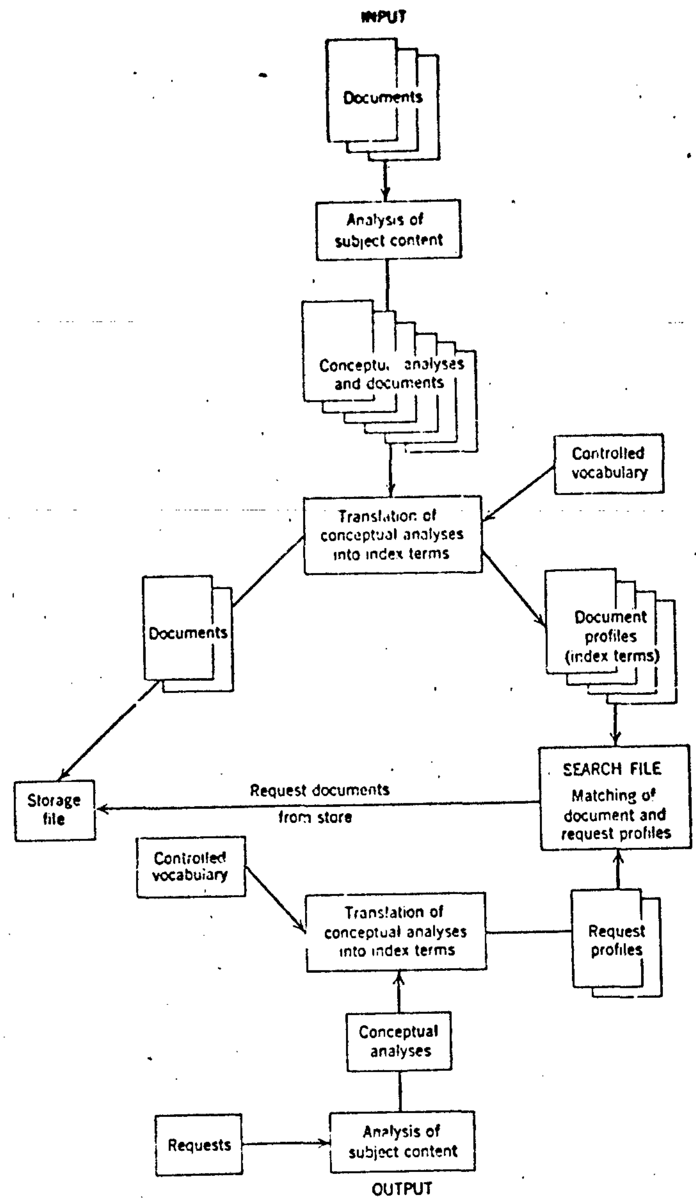
   (b) select an appropriate form of storage for source documents, either directly dependent on or independent of the search medium.

(5) Determine how to exploit the selected system by development of skillful question analysis and appropriate search strategies.

---

[2]F. Wilfrid Lancaster, Information Retrieval Systems; Characteristics, Testing, and Evaluation (New York: John Wiley & Sons, Inc., 1968), p. 4.

[3]Kent, op. cit., p. 22.



The activities of "information retrieval."

Figure 9

(6) Learn how to operate the system or cause it to operate in conducting searches.

(7) Select a means for obtaining the results of searches and copies of source documents, digests, or abstracts, or bibliographic references to them.

A. Identification and acquisition of Information

If a document is not acquired by an information system, then no matter how efficient the system is, that particular document cannot be retrieved. The extent of coverage in the subject area of interest, and the quality of the items that have been covered, are the two most important things. The system may try for extensive and comprehensive coverage in the subject area of interest or it may try to be selective and discriminating. Obviously, there will be high recall* and low precision** and vice versa in the above two situations respectively. We have to keep in mind that documents may be judged of no value for reasons like age, reliability, level or type of subject treatment, language, and so forth.

One of the most important problems in information system design is the establishment of criteria for the selection of documents. Because

---

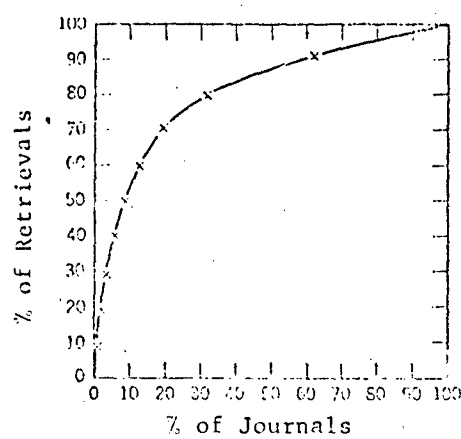
*The recall ratio is defined by the formula 100 R/C, where C is the total number of documents in the system that are established to be relevant to a particular request, and R is the number of these relevant documents that are retrieved in the conduct of a search for this request in the index to the collection.[4]

**The precision ratio is defined by the formula 100 R/L when R is the number of relevant documents retrieved in a search, and L is the total number of documents retrieved in that search.[5]

[4] Lancaster, *op. cit.*, p. 55.

[5] *Ibid.*, p. 56.

sometimes even relevant documents are judged irrelevant by the users for reasons like out of date, of doubtful validity, too mathematical, or "can't read this language." This problem typically relates to the acquisition policy. In a real-life situation it has been found that of all the articles that were retrieved in the test searches, and judged of value by requesters, approximately 90% were English. But foreign materials occupy about 40% of the data base, and are actually estimated to consume 50% of the input costs of the system. Obviously, on cost-effectiveness grounds, it is hard to justify the allocatic of 50% of input costs to 10% of total usage. In the case of journal titles also, it has been found that 10% of the journals account for about 50% of the retrievals, while 30% account for almost 80% of the retrievals,[6] as indicated in Figure 10.



PERCENTAGE OF JOURNALS ACCOUNTING FOR RETRIEVALS

Figure 10

A word of caution is in order here. We must not be too mechanical in weeding out the journals and other documents which are not earning their keep. Scientific breakthroughs do not always tread on cost-benefit

[6] Ibid., pp. 164-165.

grounds; they even have a tendency to elude averages and percentages. Fisher found a very important statistical table in an agricultural journal; and a ten-year old issue of a journal containing the report by Alexander Fleming which led to the discovery of penicilin can hardly be called "aged."

B. Analysis

When it has been decided to enter a document in the information system, it becomes a member of the universe of documents in the system. The problem is to represent the document adequately in the searchable files such as card catalog, magnetic tapes, disks, and so forth and/or to shelve the document with its like members in the document storage as the books in a library are arranged in some classified order.

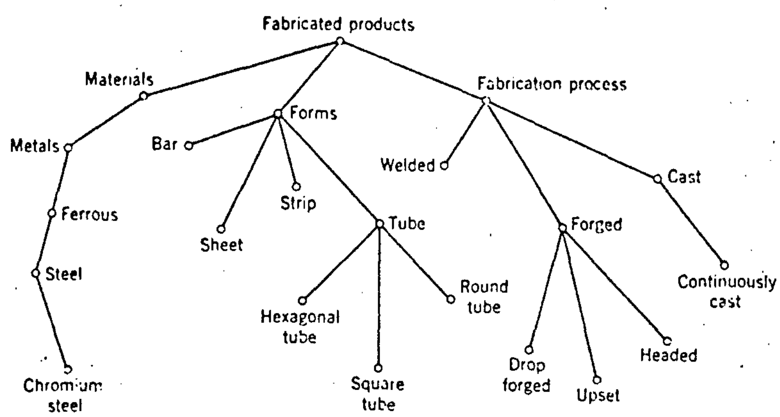
Every document or part of it belongs to one or more requests. The problem is to find the address. A request for information should get all the documents or parts thereof that are addressed to it. So the analysis of a document is a semantic problem; what it means and to which information needs it addresses itself. This is the identification of the intent of the content.

But the intent may be identified and labeled in many different ways. This is the problem of indexing by providing each document with an adequate number of direct or indirect access points, and the problem of inter-indexer inconsistencies. The process of identification involves the Boolean functions--class sum, intersection, and complementation.

In the development of indexing systems, many different ideas have been explored, such as enumeration, concept or term coordination, hierarchical thesaurus construction, KWIC and so forth. The following Figure 11 has been used by Lancaster[7] to depict how the genus "fabricated products"

[7] *Ibid*., p. 27.

may be subdivided in a classification schedule or list of subject headings. Although the hierarchy of Figure 11 enumerates, and therefore allows us to specify, such classes as "continuously cast products," "forged products," "sheet," "tube," "steel," and "chromium steel," it does not enumerate, and therefore will not allow us to specify the more complex and specific intersections of these classes, such as "continuous cast tube" or "chromium steel sheet."



THE HIERARCHICAL TREE

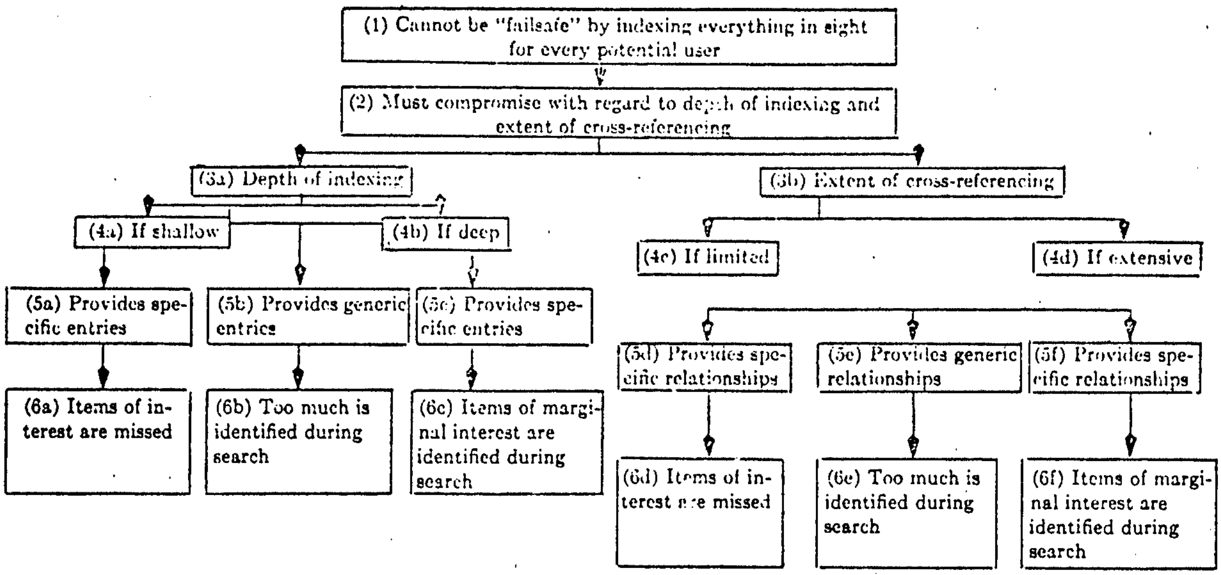
Figure 11

Deciding upon the depth of analysis and indexing is another problem. Precision and recall (page 53) have inverse and direct relationship with depth of indexing, respectively. A happy balance must be found between the noise-tolerance* propensity of the user and the depth of indexing.

---

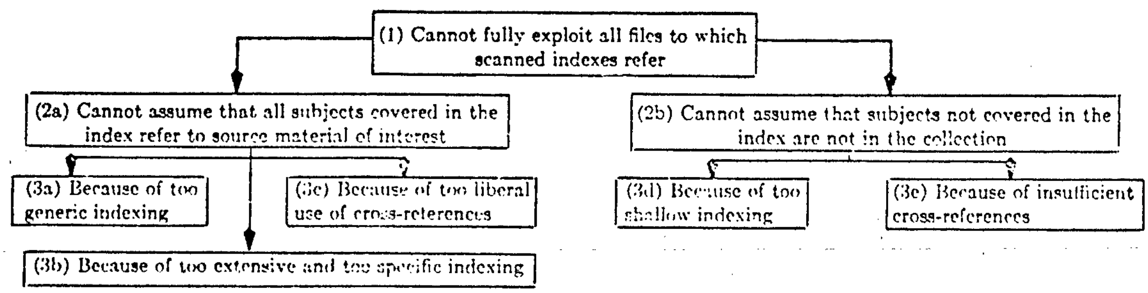
*Noise-tolerance is defined as the willingness of the user to accept a certain number of non-relevant and peripheral documents with his search output.

This dilemma has been analyzed by Kent[8] and displayed in the Figures 12 and 13.



THE DELEGEE'S DILEMMA

Figure 12



THE SEARCHER'S DILEMMA

Figure 13

---

[8] Allen Kent, Specialized Information Centers (Washington, D.C.: Spartan Books, 1965), pp. 16-17.

"We may characterize the dilemma of the delegee by considering the problems of the analyst, or more specifically, problems faced by the indexer. Basically, the dilemma revolves about the consideration that an indexer cannot determine every subject, point of view, or implication of the source materials being examined that may be of interest to *all* potential users. Economic and technical considerations prevent him from attempting to be "failsafe" in his analysis by indexing everything in sight (1). Accordingly, the first of a series of technical compromises is initiated (2) both with regard to depth of indexing (3a) and extent of cross-referencing (3b). [Figure 12].

If indexing is too shallow (4a) or cross-referencing too limited (4c) and only specific entries (5a) or specific relationships (5d) are provided, then items of interest may well be missed (6a and 6d). If only generic entries (5b) or generic relationships (5e) are provided, then too much of marginal interest may be identified during a search (6b and 6e).

On the other hand, if indexing is too deep (4b) or cross referencing too extensive (4d) and only specific entries (5c) or specific relationships (5f) are provided, then many items of only marginal interest may be identified during a search (6c and 6f). But if only generic entries (5b) or generic relationships (5e), are provided, again too much may be identified during searches of the resulting indexes (6b and 6e).

The result is a corresponding dilemma faced by the searcher [see Figure 13].

The dilemma facing the searcher of the index relates to his inability to exploit fully the files to which the indexes refer (1). It cannot be assumed that subjects that *are* covered in the index do indeed refer to source material of interest (2a) when too generic indexing (3a), too extensive and too specific indexing (3b), or too liberal use of cross-references (3c) has been used.

The other horn of the dilemma is that it cannot be assumed that subjects *not* covered in the index are *not*, nevertheless, in the collection (2b), when too shallow indexing (3d) or insufficient cross-references (3e) have been used." [9]

C. Vocabulary and Subject Heading Control

This problem is intertwined with the problem of analysis. If the system is using a controlled or restrictive vocabulary, then the indexer has to translate the analytics into the vocabulary terms which are legal in the system. A system may perform with very little control of vocabulary by using, say, the key words in a title or abstract. When the vocabulary is controlled, the lack of a term in the vocabulary may cause an indexer to either ignore the concept or use an available near term which only inadequately represents the concept concerned. The result will be recall failure in the first case and precision failure

---

[9] Ibid., p. 15.

in the second. The system should provide the indexer with tools which will help him in determining the specificity and generality of terms and how terms are subsumed under other terms.

The group concerned with the development and maintenance of the vocabulary should work in close cooperation with the indexers who apply the vocabulary to represent documents in the system. They are the right people to uncover the inadequacies in the vocabulary if there are any. The vocabulary control system should be sufficiently flexible so that it may react to the feedbacks from other system components affected by the vocabulary.

D. Recording the Results of Analysis on a Searchable Medium

This is a problem of file organization and access efficiency. The searchable file may be recorded on 3" x 5" cards, punch cards, paper or magnetic tapes, discs, and so forth. Files may be organized sequentially, record by record, or in an inverted way, aspect by aspect, with relevant document identifications following their respective aspects. Depending on the organization of the files, access may be sequential, random, binary, or a combination thereof. In a computer-based system, an optimally organized file may entail considerable cost advantages.

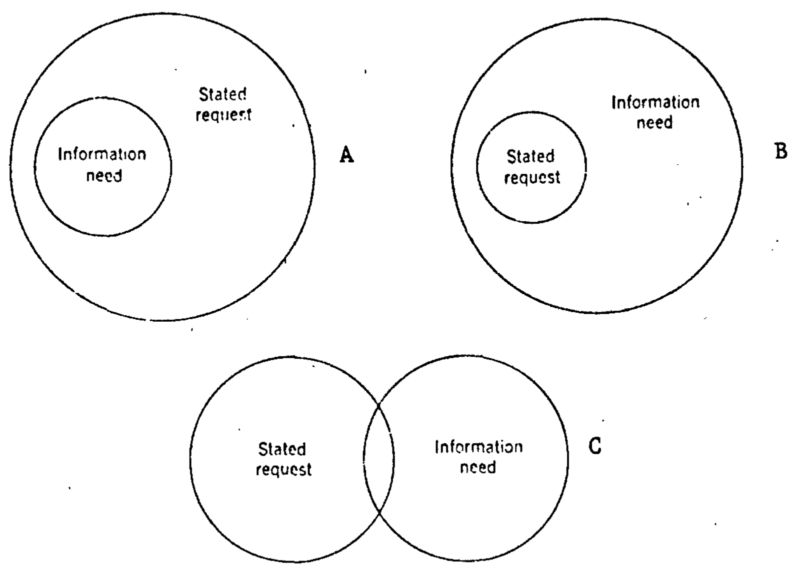
E. Storage of Records or Source Documents

Most information retrieval systems retrieve document identifiers, such as accession numbers, as search output. There are some systems which also retrieve citations, abstracts, or extracts. If the information system makes itself responsible for providing the full documents like the University

Microfilms' DATRIX services or the ERIC system, it might get involved in the problems of logistics, networking, document reproduction, microform storage, and so forth.

F. Question Analysis and Development of Search Strategy

At this point the system is interfaced with the user. This activity directly affects the search output. The analysis of documents and analysis of questions have a lot in common. Both involve the inference and identification of the intent of the author or requester, as the case may be. The following Figure 14[10] illustrates the problem of discrepancy between the stated request and the information need.



NEED-REQUEST DISCREPANCY
Figure 14

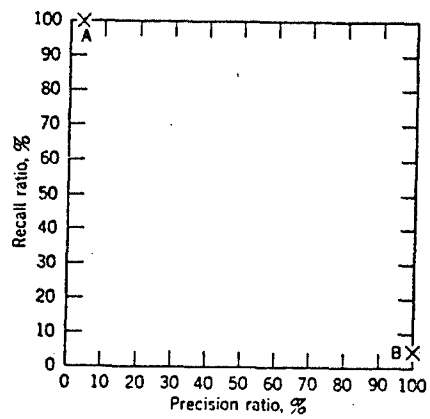
Figure 14 A indicates that the requester has asked for something much broader than his actual information need warrants; as, for example, asking for everything on ornithology when the real need is for information

---

[10]Lancaster, op. cit., pp. 146-147.

on migratory birds. Figure 14B depicts the other side of the problem, that is, the request is much too specific with respect to the information need. Figure 14C illustrates the case where there is a partial overlap between the stated request and information need.

Development of a search strategy is the process of translating the request elements into legal terms and bringing them into the desired logical relationship by using Boolean operators like alternation, intersection, and negation. It is possible to formulate a search strategy so precise (i.e., highly exhaustive and highly specific) that it would almost certainly retrieve only relevant documents, if it retrieved any at all. If the system restricts itself to such strategy formulations, it could expect to operate at 100% precision, but at a very low recall such as point B in Figure 15 below.[11]



PRECISION-RECALL TRADEOFF

Figure 15

The problem is to find out the optimum mix between precision and recall so that instead of performing either at point B of Figure 15 as explained above, or at point A where recall is very high and precision is

[11] Ibid., p. 75.

very low, the system may perform somewhere in between where the relationship between precision and recall is optimized for the system.

G. Conducting of Search

This is the process of matching the formulated search strategy with the files and retrieving the document number (in most cases) whenever there is a hit. The problem is twofold: 1) the mechanism of the search itself, and 2) the management of the search process. The mechanism of the search will depend on the organization of the files such as sequential, inverted, random, and so forth. A mistake or change of plans in this area may be very expensive. NASA Technology Transfer system changed its mind and converted its files from inverted to sequential. The process of conversion must have been expensive. Changes sometimes become necessary because of advancement in technology. Now that remote and random access capabilities are available, it may be desirable to have inverted files. The mechanism of the search will also depend on the searchable medium that has been used to record the results of analysis.

The management of the search process involves the problems associated with batch-processing, frequency of runs, queuing, time-sharing, input/output devices, and so forth. Decision making between alternatives in this area is difficult and often tied up with the system design.

H. Delivery of Results of Search

At this point the user becomes the recipient of the output against his request or profile. This is an area of user system interaction. The output is the result of system performance. So the system must make sure that it gets necessary and sufficient feedback from the user so that the system may evaluate itself. It is important to obtain a critique from the user on his relevance judgment so that we do not change or modify the system for the wrong reasons.

The search result may be delivered in the form of document citation, the full document in the original or in some form of reproduction, or the user may be provided with some surrogate* of the document such as abstracts or extracts.

If a full document is provided, there is no problem, at least for the information system, though the user may find himself inundated by the output against his request. However, the user may be given some surrogates of the retrieved documents so that he may have an opportunity of reducing the volume of output to manageable proportions by performing some relevance judgments on the basis of the surrogates.

This brings us into the problem of relevance predictability of document surrogates. An evaluation of the ability of intermediate response products (IRP's), functioning as cues to the information content of full documents, to predict the relevance determination that would be subsequently made on these documents by motivated users of information retrieval systems, was made under controlled experimental conditions by Kent.[12] The hypothesis that there might be other intermediate response products (selected extracts from the document, i.e., first paragraph, last paragraph, and the combination of first and last paragraph) that would be as representative of the full document as the traditional IRP's (citation and abstract) was tested systematically. The results showed that:

---

*Surrogate is defined as anything that can represent a document such as abstracts, summary, first paragraph, etc.

[12]Allen Kent, et al.,"Relevance Predictability in Information Retrieval Systems," Method. Inform. Med., VI, No. 2(April, 1967)45-51.

(1) There is no significant difference among the several IRP treatment groups on the number of cue evaluations of relevancy which match the subsequent user relevancy decision on the document;

(2) First and last paragraph combinations have consistently predicted relevancy to a higher degree than the other IRP's;

(3) Abstracts were undistinguished as predictors; and

(4) The apparent high predictability rating for citations was not substantive.

The desideratum here is to be able to give the user that particular surrogate mix of the output which would enable him to predict the relevancy of the documents concerned with maximum probability of success.

We have now discussed the problems of the information system in general. We now know its properties, characteristics, and objectives. So, at this point we may relate all this to a real-life, operational information system, eventually to find out how the design of a system affects its performance and survival potential. The Medical Literature Analysis and Retrieval System (MEDLARS) of the National Library of Medicine is a large-scale, computer-based information system. It is an operational system, now entering its second phase, and has recently gone through an evaluation study, revealing some important facts which have been heavily used in this dissertation. From these facts it appears that the problems faced by MEDLARS are generically peculiar to all information systems, and this makes MEDLARS an ideal object system for our purpose. The next chapter deals with MEDLARS.[13]

---

[13]All factual data and the system flow chart of MEDLARS came from Charles J. Austin, MEDLARS 1963-1967, (Bethesda: National Library of Medicine, 1969).

VI. PROBLEM OF INFORMATION SYSTEM: MEDLARS

A. Background and History

The Library, now known as the National Library of Medicine (NLM), initiated its program of bibliographic control of the medical literature in 1879 with the publication of the first issue of Index Medicus, which continued until 1927. Replaced from 1927 to 1956 by the Quarterly Cumulative Index Medicus, published by the American Medical Association, Index Medicus reappeared as an NLM publication in 1960, replacing the monthly Current List of Medical Literature.

Index Medicus was produced by a partially mechanized system known as the Listomatic System, from 1960 to 1963, which aided in the subsequent development of MEDLARS in the following way:

(1) Provided much background data used in the design of MEDLARS;

(2) Offered a valuable operating experience on which to base the system design; and

(3) Assisted in the data conversion task for MEDLARS.

The Listomatic Camera System worked effectively in the publication of Index Medicus and related publications; however, it had very limited information retrieval capability.

The rapidly growing size of Index Medicus and the limitations of the Listomatic system caused the NLM to start planning a new and more highly mechanized system. Ultimately a contract was awarded to the General Electric Company, Information Systems Operation, Bethesda, Maryland. The conversion period ran from April to December 1963. Approximately 45,000 journal article citations from the 1963 Index Medicus were converted to magnetic tape. Cut-over to the new system was accomplished in January 1964, and it has been in operation continuously since that date.

B. System Objectives

The major objectives for the MEDLARS system as stated by the NLM Management in 1961 are as follows:

(1) Improve the quality of and enlarge (broaden the scope of) Index Medicus and at the same time reduce the time required to prepare the monthly edition for printing from 22 to 5 working days.

(2) Make possible the production of other compilations similar to Index Medicus in form and content (but in more specific medical subject areas and hence smaller in size).

(3) Make possible, for Index Medicus and other compilations, the inclusion of citations derived from other sources as well as from journal articles.

(4) Make possible the prompt (a maximum of two days) and efficient servicing of requests for special bibliographies, on both a demand and a recurring basis, regularly searching up to five years of stored computer files.

(5) Increase the average depth of indexing per article (number of descriptive subject terms per article) by a factor of five, i.e., ten headings versus two.

(6) Nearly double the number of articles that may be handled (indexed and entered into the computer) annually--from 140,000 now to 250,000 in 1969.

(7) Reduce the need for duplicative total literature screening operations (at other libraries and information centers).

(8) Keep statistics and perform analyses of its own operations to provide the information needed to monitor and improve system effectiveness.

(9) Permit future expansion to incorporate new and as yet not completely defined--and hence secondary--objectives.

MEDLARS is not a newly developed system; it grew out of an existing system which was operating inadequately as we have seen. If we study the major objectives laid out in 1961 for MEDLARS, we can easily see that most of these objectives were conceived as corrective measures with some augmentation of the existing system. It may be interesting to see how these objectives fit in the unit operations format as laid out on pages 49-50. In this discussion Index Medicus and the Recurring Bibliographies will be considered as search outputs.

Objective No. 1 relating to quality, scope and speed of Index Medicus publication may be considered as belong to the Unit Operations: Acquisition, Conducting of Search, and Delivery of Results of Search.

Objective No. 2 belongs to Unit Operations: Conducting of Search (6), and Delivery of Results of Search (7) since it relates to the production of special compilations.

Objective No. 3 again involves the Unit Operation Acquisition because it aims at the inclusion of citations derived from "other" sources.

Objective No. 4 relates to prompt and efficient servicing of requests, hence it should come under the Unit Operations No. 6 and 7; that is, Conducting of Search, and Delivery of Results of Search.

Objective No. 5 is concerned with depth of indexing and should go under the Unit Operations No. 1, Analysis and No. 2, Vocabulary and Subject Heading Control.

Objective No. 6 involves three of the Unit Operations; namely, Analysis (1), Vocabulary and Subject Heading Control (2), and Recording of Results of Analysis on a searchable medium (3), because it intends to nearly double the number of articles that may be indexed and entered into the computer annually.

Objective No. 7 intends to make MEDLARS good enough as a one-stop information system capable of reducing the need for duplicative total literature screening operations at other libraries and information centers. So this obviously relates to the unit operation--Acquisition.

Objectives No. 8 and No. 9 relate to system evaluation, and flexibility and growth. Kent did not consider these as unit operations. It is not difficult, however, to consider R & D as an integral part of each and every unit operation.

The following Table 2 shows the distribution of the objectives into the unit operations. It is interesting to note that none of the objectives relates to the unit operations: Storage of Records or Source Documents (4), and Question Analysis and Development of Search Strategy (5). Concentration of checks (x) would indicate the main problem areas.

Objectives \ Unit op.	A*	1	2	3	4	5	6	7	
1	x						x	x	
2							x	x	
3	x								
4							x	x	
5		x	x						
6		x	x	x					
7	x								
8									Do not apply
9									Do not apply

*A = Acquisition

DISTRIBUTION OF THE OBJECTIVES
INTO THE UNIT OPERATIONS

TABLE 2

C. Design Criteria

Some of the major guiding principles on which the design of MEDLARS was based are as follows:

First was a decision to continue to use human indexers for assigning subject descriptors to the literature for subsequent retrieval and publication of references. (The state of the art of automatic indexing in 1961 was such that it was not considered feasible for MEDLARS.) A second decision was to continue to use a controlled vocabulary for indexing.

Another major decision was to index each article only once, and use a single computer input record both for publication in Index Medicus and for retrieval purposes.

Other important design criteria included:

(1) A decision to train search specialists for formulating retrieval requests for the computer, rather than allow customers of the system to attempt to formulate their own computer search statements.

(2) A decision to use serial magnetic tape files for storing journal article citations, rather than random access devices. (This was also a decision influenced by the 1961-62 state of the art.)

(3) A decision to segment computer programs into self-contained "modules" for ease of maintenance and system changes.

(4) A requirement that the system employ a "high-quality" output device, superior to available computer printers for preparation of copy for MEDLARS publications.

(5) A decision not to increase the amount of clerical work required of the professional indexers by using clerical personnel for

preparation of the computer input record. It was decided also to design the system so as to use the computer for as much coding and editing of the input data as possible.

We may now look into the question of how the design criteria established for MEDLARS relate to its objectives.

The use of human indexers and controlled vocabulary relate to the objective to increase the average depth of indexing (No. 5). These are tied up with the availability of technology. The decision to use a single computer input record for multiple use is going to help several of the objectives, such as compilations similar to Index Medicus (No. 2) and efficient servicing of requests for special bibliographies (No. 4).

Formulation of search strategies by system personnel (criterion 1) is going to help realize the objective of efficient servicing of requests from the users (No. 4). The decision to use serial files rather than random access devices (criterion 2) is contingent upon the availability of technology and compatible with non-remote access environment (relates to No. 4 since it concerns file access).

The design criterion 3 regarding modular approach relates to the objective of future expansion and incorporation of new objectives (9).

The requirement of a "high-quality" output device (criterion 4) is related to the rapid publication of Index Medicus and other compilations (1). It is interesting to note that the system did not want to restrict itself on this count by the available technology.

The criterion of using clerical personnel for preparation of the computer input record, and using the computer for as much coding and editing of the input data as possible (criterion 5), relates to the objectives of doubling the number of articles that may be indexed and entered into the computer annually (6).

The following Table 3 relates the objectives with the Design Criteria:

TABLE 3

OBJECTIVES/DESIGN CRITERIA RELATIONSHIP

Design Criteria \ Objectives	1	2	3	4	5	6	7	8	9
Use of Human Indexers					x				
Single Input Record		x		x					
1				x					
2				x					
3									x
4	x								
5						x			

D. System Description

The products of MEDLARS can be divided into two major categories: 1) bibliographic publications designed for use by a large group of people working in related fields; and 2) individual demand searches of the literature tailored to the stated requirements of an individual or small group of people working on the same project. Demand searches considered to be of broader interest to people other than the person originating the search request are reprinted as "Literature Searches" and copies are sent to anyone upon request. In addition to publications and demand searches, MEDLARS also produces internal reports to be used by operating and management personnel. The data flow through MEDLARS is represented in the flow chart in Figure 16.

MEDLARS can be functionally divided into three major parts:

(1) Input Subsystem

(2) Retrieval Subsystem

(3) Publication Subsystem

The Input Subsystem is a man-machine interface where the intellectual work of the literature analyst is combined with the processing

MEDLARS SYSTEM OVERALL DATA FLOW CHART *

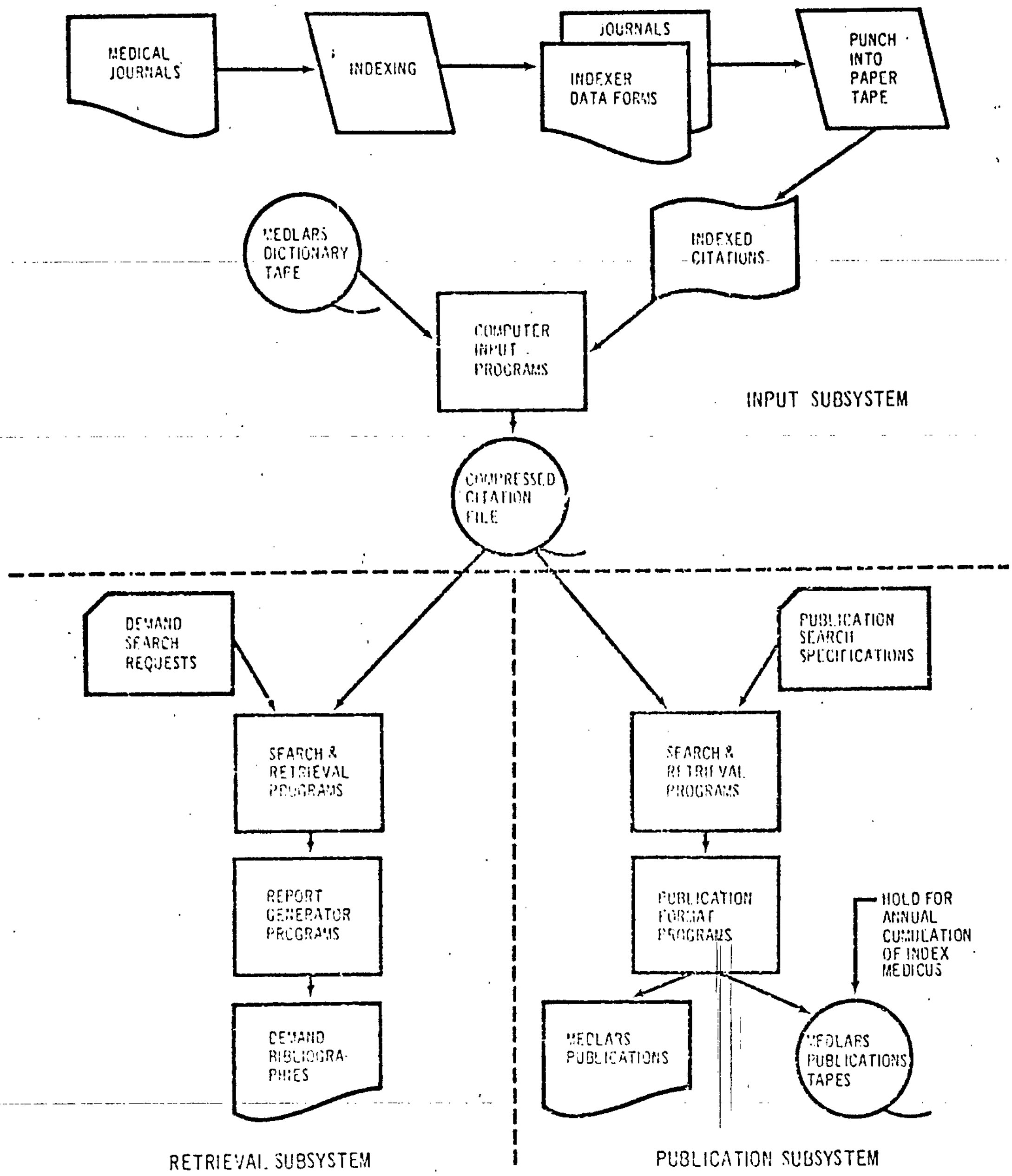


Figure 16

*Source: Austin, op. cit., p. 10.

and storage capabilities of the computer.

The Retrieval Subsystem handles the requests for demand bibliographies. Search specialists formulate the request into a list of search parameters linked in logical fashion. The formulated search requests are punched into cards and batched for daily computer processing. The search and retrieval programs match a batch of search questions against every record in the Compressed Citation File. Citations retrieved are printed in any one of a variety of output formats by means of print programs.

The Publication Subsystem is concerned with the preparation of periodic indexes to current biomedical literature. In accordance with a publication schedule, search specification cards are entered into the computer for bibliographies to be compiled. The search and retrieval programs retrieve the appropriate citations from the Compressed Citation File. The Photon 900 computer phototypesetter--Grace--is used in the process of printing the final publication.

E. MEDLARS Evaluation

In January 1966, the National Library of Medicine embarked upon the detailed planning of a test program to evaluate the performance of MEDLARS. In December 1965, Mr. F. W. Lancaster was recruited by the Library to fill the new position of Information Systems Evaluator so that the evaluation could be conducted in a completely impartial manner by some one who had in no way been concerned with either the design or operation of the MEDLARS system. In addition, a MEDLARS Evaluation Advisory Committee was formed to review the design and execution of the test program, and the analysis and presentation of the test results. Cyril W. Cleverdon, Librarian, College of Aeronautics, Cranfield, England, served as a special consultant to the Library on the Evaluation Project.

The Evaluation Project studied the performance of MEDLARS in relation to 300 actual requests made to the system in 1966 and 1967. This is the first large-scale evaluation of a major operating information system. Dr. Martin M. Cummings, Director of the National Library of Medicine, emphasized that to remain responsive to the demands of its users, a large scientific or technical information system must examine itself critically, and hoped that a major benefit of this investigation will be the establishment of a program for the continuous quality control of MEDLARS products and services.[1]

F. Objectives of the Test Program[2]

The principal objectives of the test program may be summarized as follows:

(1) To study the demand search requirements of MEDLARS users.

(2) To determine how effectively and efficiently the present MEDLARS service is meeting these requirements.

(3) To recognize factors adversely affecting the performance of MEDLARS.

(4) To disclose ways in which the requirements of MEDLARS users may be satisfied more efficiently and/or more economically. In particular, to suggest means whereby new generations of

[1]F. Wilfrid Lancaster, Evaluation of the MEDLARS Demand Search Service (Washington, D. C.: U. S. Department of Health, Education, and Welfare, Jan. 1968), Preface, p. iii.

[2]Ibid., pp. 8-10.

equipment and programs may be used most effectively in in satisfaction of demand search requirements.

In addition, the test was expected to produce further valuable benefits:

(5) On the basis of test results, and analyses of failures, it would aid in establishing methods that could be used to implement a continuous "quality control" program for the MEDLARS operation.

(6) The test would provide a corpus (of documents, requests, indexing, search formulations, and "relevance" assessments) that could be used for further tests and experimentation.

(7) It would identify specialized areas that might require further experimentation and evaluation.

G. Test Requirements

It is assumed that the prime requirements of demand search users relate to the following factors:

(1) The <u>coverage</u> of MEDLARS (i.e., the proportion of the useful literature on a particular topic, within the time limits imposed, that is indexed into the system).

(2) Its <u>recall</u> power (i.e., its ability to retrieve "relevant" documents, which, within the context of this evaluation, means documents of value in relation to an information need that prompted a request to MEDLARS).

(3) Its <u>precision</u> power (i.e., its ability to hold back "non-relevant" documents).

(4) The response time of the system (i.e., the time elapsing between receipt of a request at a MEDLARS center and delivery to the user of a printed bibliography).

(5) The format in which search results are presented.

(6) The amount of effort the user must personally expend in order to achieve a satisfactory response from the system.

It follows, therefore, that the test had to establish user requirements and tolerances in relation to these various factors.

In particular, the test was designed to answer certain specific questions relating to the operating efficiency of the MEDLARS demand search service. These questions are enumerated below:

(1) Overall performance

a. What is the overall performance level of MEDLARS in relation to user requirements? Are there significant differences for various types of requests and in various broad subject areas?

(2) Coverage and processing

a. How sound are present policies regarding indexing coverage?

b. Is the delay between the receipt of a journal and its appearance in the indexing system significantly affecting performance?

(3) Indexing

a. Are there significant variations in inter-indexer performance?

b. How far is this related to experience in indexing and to degree of "revising?"

c. Do the indexers recognize the specific concepts that are of interest to various user groups?

d. What is the effect of present policies relating to exhaustivity of indexing? In particular, is there a significant difference between retrieval performance for articles from "depth-indexed" and "non-depth-indexed" journals? What would be the effect of searching on only Index Medicus headings?

(4) Index language

a. Are the terms sufficiently specific?

b. Are variations in specificity of terms in different areas significantly affecting performance?

c. Are pre-coordinate* type terms and subheadings, which have been included to meet the requirements of Index Medicus, hindering the efficiency of retrieval by MEDLARS?

d. Is the need for additional precision devices, such as weighting, role indicators, or a form of interlocking, indicated?

e. Is the quality of term association in MeSH satisfactory?

f. Is the present "entry vocabulary" adequate?

(5) Searching

a. What are the requirements of the users regarding recall and precision?

---

*Pre-coordinate system: System in which class relationships are expressed once and for all, by the labels used to define classes in the indexing operation is called pre-coordinate system, e.g., Labor Economics.[3]

[3]Lancaster, Information Retrieval Systems, pp. 33-34.

b. Can search strategies be devised to meet requirements for high recall or high precision?

c. How effectively can NLM searchers screen output? What effect does screening have on recall and precision figures?

d. What are the most promising modes of user/system interaction?

   1) Having more liaison with information staff at the local level?

   2) Having more liaison directly with MEDLARS search analysts?

   3) Certain alternative modes of interaction (e.g., user examination of proposed search strategy, or iterative search) not presently used in the MEDLARS operation?

e. What is the effect on response time of these various modes of interaction?

f. Are there significant differences in performance between the various MEDLARS centers?

(6) Input and computer processing

a. Do input and data processing procedures, including various clerical functions, result in a significant number of search failures?

VII. ANALYSIS OF THE RESULTS OF THE TEST PROGRAM

We have now studied MEDLARS and related its system objectives to the Unit Operations. We have also noted the objectives of the Test Program. Now we are ready for the analysis of the results of the Test Program. It is not feasible or necessary, for our purpose, however, to study the Lancaster Evaluation in all its aspects. So we have concentrated our attention to one of the functions--the subject indexing function. We are repeating the table of unit operations here to facilitate reference:

(1) Analysis, involving perusal of the record and the selection of points of view (or analytics) that are considered to be of sufficient probable importance to warrant the effort of rendering them searchable in the system.

(2) Vocabulary and subject heading control, involving establishment of some arbitrary relationships among analytics in the system. These arbitrary relationships are usually dependent on similarities among analytics as revealed in dictionary definitions for the words used to express the analytics.

(3) Recording of results of analysis on a searchable medium, involving the use of a card, tape, film, or other medium, on which the analytics are transcribed.

(4) Storage of records, or source documents, involving the physical placement of the record in some location, either in its original form, or transcribed or copied (in full or reduced size) onto a new medium.

(5) Question analysis and development of search strategy, involving the expression of a question or a problem, the

selection of analytics based on analysis of the question, the expression of these analytics in terms of a particular search mechanism, and their arrangement into a configuration that represents a probable link between the question as expressed and the records on file as analyzed.

(6) Conducting of search, involving the manipulation or operation of the search mechanism in order to identify records from the file.

(7) Delivery of results of search, involving the physical removal or copying of a record from file in order to provide it in response to a request.

The subject indexing function of MEDLARS relating to the unit operations of analysis, vocabulary control, and search strategy formulation has been selected for intensive analysis and application of the design methodology because proper operation of this function is probably the most important single factor governing the performance of an information retrieval system. As Lancaster has pointed out in his MEDLARS Evaluation, on which the following discussion is based, "Poor searching strategies, and inadequate or inconsistent indexing, can mar the performance of a system, but indexing and searching, however good, cannot compensate for an inadequate index language. In other words indexers and searchers can perform only as well as the index language allows."[1]

An analysis of the reasons for the MEDLARS demand search failures shows that almost all of the failures can be attributed to some aspect of indexing, searching, the index language (i.e., MeSH and its auxiliaries),

---

[1]Lancaster, Evaluation of the MEDLARS Demand Search Service, p. 80.

computer processing, or the area of interaction between the requester and the system. The Lancaster study isolated a single "most critical" cause for any one failure, wherever possible.

The principal objectives of the Lancaster Evaluation study, so far as indexing is concerned, are to answer the following questions.

A. Indexing

(1) Are there significant variations in inter-indexer performance?

(2) How far is this related to experience in indexing, and to degree of "revising"

(3) Do the indexers recognize the specific concepts that are of interest to various user groups?

(4) What is the effect of present policies relating to exhaustivity of indexing? In particular, is there a significant difference between retrieval performance for articles from "depth-indexed" and "non-depth-indexed" journals? What would be the effect of searching on only Index Medicus headings?

B. Index Language

(1) Are the terms sufficiently specific?

(2) Are variations in specificity of terms in different areas significantly affecting performance?

(3) Are pre-coordinate type terms and subheadings, which have been included to meet the requirements of Index Medicus, hindering the efficiency of retrieval by MEDLARS?

(4) Is the need for additional precision devices, such as weighting, role indicators, or a form of interlocking, indicated?

(5) Is the quality of term association in MeSH satisfactory?

(6) Is the present "entry vocabulary" adequate?

C. Exhaustivity, Specificity, and Entry Vocabulary
(Unit operation 2)

In the analyses of failures three terms, namely, exhaustivity, specificity, and entry vocabulary, have been used with special meaning. By exhaustivity of indexing is meant the extent to which the potentially indexable items of subject matter contained in a document are in fact recognized in the "conceptual analysis" stage of indexing and translated into the language of the system. A high level of exhaustivity of indexing will tend to result in a high recall performance for a retrieval system, but also in a low precision performance. Conversely, a low level of exhaustivity of indexing (i.e., inclusion of "most important" concepts only) will tend to produce a high precision, low recall performance. Exhaustivity of indexing is largely controlled by a policy decision of system management. Failure to retrieve a relevant document due to the fact that a particular concept was not indexed is called a recall failure, and the retrieval of an unwanted document because of inclusion of minor importance concepts in indexing is called a precision failure due to exhaustivity of indexing.

Specificity of indexing refers to the generic level at which a particular item of subject matter is recognized in indexing. For example, the topic "tetrodotoxin" could be expressed specifically by a single term TETRODOTOXIN, or a decision could be made to express this subject precisely by the joint use of two terms, TOXINS and PUFFER FISH, and recording this decision in the MEDLARS entry vocabulary as: Tetrodotoxin <u>index under</u> TOXINS and PUFFER FISH. From the point of view of recall, it matters little whether a class is uniquely defined or subsummed under some larger class, as long as the decision taken is recorded in the entry vocabulary.

The following Tables 4 and 5 show that the indexing subsystem contributed to 37% of the recall failures, and was in fact the largest contributor to this group of failures, but to only 13% of the precision failures.[2]

D. Types of Indexing Failures
(Unit operation 2)

There have been two distinct types of indexing failure:

(1) Those due to indexer errors; and

(2) Those due to a policy decision governing the number of terms assigned to an article (i.e., the policy regarding exhaustivity of indexing).

Indexer errors are themselves of two types: 1) omission of a term or terms necessary to describe an important topic discussed in an article, and 2) use of a term that appears inappropriate to the subject matter of the article. Omission will normally lead to recall failures, while use of an inappropriate term can cause either a precision failure (the searcher uses this term in a strategy and retrieves an irrelevant item) or a recall failure (the searcher uses the correct terms and a wanted document is missed because labeled with an incorrect term). The reason for the use of inappropriate terms appears to be the general misuse of a particular term at some point in time. Lancaster gives the example, RADIOISOTOPE SCANNING which has been used indiscriminately

---

[2] _Ibid_., p. 49.

TABLE 4

REASONS FOR 797 RECALL FAILURES

(302 searches were examined; and in 238 of these recall failures are known to have occurred).

Source of Failure	Number of Missed Articles Involved	Percentage of Total Recall Failures Involved	Number of Searches Involved	Percentage of the 238 Searches Involved
Index Language				
Lack of appropriate specific terms	81	10.2%	29	12.2%
Indexing				
Insufficiently specific	46	5.8%	31	13.0%
Insufficiently exhaustive	162	20.3%	100	42.0%
Exhaustive indexing (searches involving negations)	5	0.6%	4	1.7%
Indexer omitted important concept	78	9.8%	61	25.6%
Indexer used inappropriate term	7	0.97%	7	2.9%
TOTAL FAILURES ATTRIBUTED TO INDEXING	298	37.4%	203	85.3%

TABLE 5

REASONS FOR 3038 PRECISION FAILURES

(302 searches were examined, and in 278 of these precision failures are known to have occurred).

Source of Failure	Number of Unwanted Articles Involved	Percentage of Total Precision Failures	Number of Searches Involved	Percentage of the 278 Searches Involved
Index Language				
Lack of appropriate specific terms	534	17.6%	58	20.9%
False coordinations	344	11.3%	108	38.8%
Incorrect term relationships	207	6.8%	84	30.2%
Defect in hierarchical structure	9	0.3%	5	1.8%
TOTAL FAILURES ATTRIBUTED TO INDEX LANGUAGE	1094	36.0%	255	91.7%
Indexing				
Exhaustive indexing	350	11.5%	137	49.3%
Insufficiently exhaustive (searches involving negations)	5	0.2%	2	0.7%
Indexer omitted important concept (search involving negations)	1	0.03%	1	0.4%
Insufficiently specific	1	0.03%	1	0.4%
Indexer used inappropriate term	36	1.2%	26	9.4%
TOTAL FAILURES ATTRIBUTED TO INDEXING	393	12.9%	167	60.1%

for any radioisotope monitoring operation, whether or not scanning was involved.[3]

"A significant number of . . . cases of indexer omissions can be attributed to the fact that no MeSH term exists for the missed notion, and there is nothing in the entry vocabulary to say how the topic is to be indexed. As a result, the indexer either omits the topic entirely or indexes it much too generally." Lancaster gives an example of a major value article to a question, unretrieved, dealing with flavin photodeiodination of thyroxine. "There is no MeSH term for 'photodeiodination,' or indeed for 'deiodination,' and there is nothing in the entry vocabulary to say how this concept is to be indexed. Consequently, the notion was completely ignored in indexing, although it might reasonably have been translated into IODINE."[4]

In the operation of any retrieval system, there will be recall failures caused by indexing that is not sufficiently exhaustive, and there will be precision failures due primarily to the fact that exhaustive indexing has brought out documents on topics for which they contain very little information. In MEDLARS this phenomenon gets compounded because of "depth" and "non-depth" treatment of journals.

Twenty percent of the recall failures are attributed to lack of exhaustivity of indexing, while 11.5% of the precision failures are caused largely by exhaustive indexing. Since September 1964, the complete list of journals indexed has been divided into two parts: "depth" and "non-depth." Articles from "depth" journals (about one

---

[3] Ibid.

[4] Ibid., p. 51.

third of all the 2400 journals regularly indexed) are presently indexed at an average of about ten index terms per article, while the non-depth articles are indexed at an average of slightly less than four terms per article.

Some of the terms assigned to both depth and non-depth articles are chosen to be the headings under which entries for the articles will appear in Index Medicus. Only the terms representing the most important topics discussed in an article are chosen as "print" or IM (Index Medicus) terms. Thus, the "print" terms can function as weighted index terms.

E. Little Use of "Weighting" (Unit operation 5 and 6)

"The author was surprised to discover, throughout the search analyses, that very little use was made of 'weighting' as a retrieval device, although MEDLARS has a built-in term weighting system in the distinction between print and non-print terms. In less than 5% of all the test searches was use made of 'print' terms to improve the precision of a search."[5] By weighting index terms, much of the irrelevant material brought out by exhaustive indexing could be screened out.

In MEDLARS, lack of specificity and lack of exhaustivity of indexing are both closely related to policy regarding indexing depth (i.e., the average number of terms assigned). Articles from non-depth journals tend to be indexed in general terms. For example, a search on spina bifida and anencephalus failed to retrieve a number of non-depth articles because they were indexed more generally under ABNORMALITIES.

[5] Ibid., p. 74.

In depth indexing, the specific malformations would have been indexed.[6]

The artificial separation of all MEDLARS journals into depth and non-depth appears, from the detailed search analyses, to lead to indexing anomalies that can cause both recall and precision failures. Although many of the articles from non-depth journals seem somewhat superficial and repetitive, others are very substantial papers which, because of a general policy decision, are indexed completely inadequately. On the other hand, half-column letters in Lancet are sometimes assigned 15-20 terms, and are thus retrieved in searches to which they contribute little or nothing. A policy of treating each article on its own merit, whatever journal it comes from, should reduce such seeming anomalies.

The indexing policy with regard to review articles appears to be particularly suspect. Review articles are indexed "non-depth" on the grounds that the material reviewed "was probably indexed in depth in the original." This is hard to justify on a number of grounds:

(1) Some of the "reviewed" literature predates MEDLARS;

(2) A good reviewer may present data in new relationships not revealed by the original articles; and

(3) A review article may contain one of the most substantial discussions anywhere in existance on a comparatively rare subject.

From the point of view of machine retrieval, the policy of indexing non-depth articles in general terms is indefensible. To quote but one example, in the analysis of a search, an article from a

[6] Ibid., p. 59.

non-depth journal (Poultry Science) entitled 'Role of streptococcus faecalis in the antibiotic growth effect in chickens' was examined. Found by manual search, but missed by MEDLARS, it was indexed only under EXPERIMENTAL LAB STUDY, INTESTINAL MICROORGANISMS and POULTRY.

Use of the general term INTESTINAL MICROORGANISM for the specific organism implicated is inexcusable. On the basis of this indexing, one could not reasonably expect the article to be retrieved in response to a request on "streptococcus faecalis in poultry" or one on "effect of penicillin on streptococcus faecalis" or even one on "antibiotic growth effect in poultry" to all of which specific topics it is highly relevant. In fact, on the basis of the indexing, one could only reasonably expect to retrieve it in a search on intestinal microorganisms of poultry, to which general subject it is indeed a slight contribution.

It is always a mistake to index specific topics under general terms. In the above example, use of the term STREPTOCOCCUS FAECALIS would allow retrieval of this item in response to a request involving this precise organism. On the other hand, the article could still be retrieved in a more general search relating to intestinal microorganisms, because the searcher is able to "explode" on all bacteria terms. The article could have been indexed very adequately under five terms: POULTRY, PENICILLIN, STREPTOCOCCUS FAECALIS, GROWTH, and EXPERIMENTAL LAB STUDY. As presently indexed, it is difficult to visualize a single retrospective search in which it would be retrieved and judged of major value. In other words, this citation and others indexed in such general terms are merely occupying space on the citation file. [7] "The present division

[7] Ibid., pp. 60-62.

of journals into 'depth' and 'non-depth' has led to indexing anomalies and to the situation in which non-depth articles occupy 45% of the file but account for only 25% of the retrievals; some of the non-depth articles are never likely to be retrieved and judged of value because they are indexed much too generally."[8]

F. Terms Omitted or Changed
(Unit operation 2)

It is difficult to evaluate the components of the overall input subsystem in MEDLARS. There appears to be no guarantee that the terms on the citation file are actually the terms assigned by the indexers. Some terms, for example, could be omitted or changed in the computer input (flexowriter) operations; others could be lost through imperfect file maintenance procedures. On the basis of a test, Lancaster has been forced to conclude that perhaps 25% of the failures attributed to indexer omissions in fact occurred later than the indexing stage. One of the test cases shows that a term (PARATHYROID GLANDS) was included on the indexer data sheet and was also included on the flexowriter proof copy. The term also appeared with the citation in the December 1966 issue of Index Medicus. "The fact that a citation printout now reveals that this term (PARATHYROID GLANDS) is no longer carried among the tracings for the article, indicates some subsequent failure of file maintenance procedures."[9]

---

[8] Ibid., p. 199.

[9] Ibid., p. 62.

G. Entry Vocabulary Should Tell Where to Look (Unit operation 2)

It has been said before that the quality of index language is probably the most important single factor governing the performance of a retrieval system. To return to the earlier discussion on the matter of the entry vocabulary, even though the class "tetrodotoxin" is not uniquely defined, it must be included in the entry vocabulary as a reference:

Tetrodotoxin <u>use</u> ANIMAL TOXINS <u>and</u> FISH

It would be done to:

(1) Indicate that documents on this specific topic have been input to the system;

(2) Ensure that all indexers use the same term combination to enter into the system articles on this precise topic; and

(3) Ensure that searchers use the right term combination to retrieve relevant literature on this topic.

Thus, although the class "tetrodotoxin" is not uniquely defined, literature on this precise topic will still be retrieved because the entry vocabulary tells precisely where to look. In this case, lack of specificity in the vocabulary will not cause recall failures. It is true articles on tetrodotoxin alone cannot be retrieved and this will mean precision failures in a search on tetrodotoxin. In other words, if a particular class of documents is not uniquely defined, but indicated in the entry vocabulary how the class has been subsumed, there will be precision failures due to lack of specificity in the vocabulary, but no recall failures will be attributable to the cause. However, if the notion is omitted even from the entry vocabulary, we will get both recall and precision failures.

H. Weakness of the Indexing Language (Unit operation 2)

To use the MEDLARS indexing language, "Acute Cecitis" must be translated into either CECAL DISEASES, or CECUM and INFLAMMATION. These retrieved 121 citations, of which but a handful were relevant, and achieved only 33.3% recall.[10] This is only an example illustrating the overall index language deficiencies in MEDLARS.

The system is particularly weak in some areas. The behavioral sciences is an example. In the area of "technics," 27.6% of all the searches are affected by lack of specificity. A quarter of the PHYSICS/BIOLOGY searches are affected by lack of specificity in the vocabulary (it is difficult to distinguish various types of radiation; e.g., ionizing from non-ionizing).

On the whole the search analyses by Lancaster have shown the MEDLARS vocabulary to be unexpectedly weak in the clinical area. Not only does it fail to express precisely a significant proportion of the pathological conditions occurring in requests, some of which are not particularly obscure (e.g., perforation of the gall bladder), but it is also deficient in its ability to express various characteristics of a disease. For example, the extent of pathological involvement cannot be indicated. Nor can we readily distinguish: acute from chronic; versions of a disease according to etiology (e.g., bacterial from non-bacterial asthma); symptomatic from asymptomatic; co-existent, unrelated conditions from true sequelae; or the situation of one disease "masquerading" as another. Again from the search analyses, the vocabulary appears weak

---

[10] Ibid., p. 85.

in areas that impinge upon medicine.[11] Introduction of subheadings, in 1966, markedly increased the specificity of the vocabulary. It is now possible to express various notions (e.g., "epidemiology" and "etiology") which were not adequately covered in the vocabulary before the subheadings were introduced. Nevertheless, it is difficult to understand why the subheadings were dropped in the first place.

I. Lancaster's General Observations on the MEDLARS Index Language

(1) There are certain types of requests being made of MEDLARS which are attempted, but with which the vocabulary is completely unable to cope, such as osteomyelitis of unknown etiology.

(2) Even with tree structures, the vocabulary is not as helpful as it could be to indexers and searchers. It is difficult sometimes to think of all terms that are possibly related to a request. Further relationships, built into the hierarchical displays, could be of great assistance to the searcher, and might well help to reduce those recall failures attributed to the searcher not covering all reasonable approaches to retrieval.

(3) Methods presently used to update the MEDLARS vocabulary are not optimally responsive to the requirements of the demand search function. Heavy reliance is placed on committees of subject specialists to review terminology in particular areas. The use of such committees tends, of course, to ensure that MeSH reflects current medical terminology. This may be highly desirable for the published bibliography, *Index Medicus*, but is not necessarily the principal requirement for vocabulary development

---

[11] *Ibid.*, p. 87

in a retrospective search system based on the coordination of terms at the time of searching.

J. No Routine Procedures to Correct Vocabulary Inadequacies (Unit operation 2)

A vocabulary tends to be most responsive when it has a high degree of literary warrant. In other words, the most valuable raw materials for vocabulary development are incoming articles and, crucial, requests being made to the system. Yet these are the very materials that appear most neglected in the development of the MEDLARS index language. Within the evaluation program, requests have been systematically analyzed from the point of view of the capability of the vocabulary to cope with them, but this is not done as part of the regular operations of the system (unit operation 5). Although a form (Request for Medical Subject-Heading Change) is available to record suggestions of indexers and searchers, very little use appears to be made of this. In other words, there are no routine, established procedures whereby indexers and searchers are required to notify the MeSH group whenever they discover either 1) an article on a topic that cannot adequately be covered in indexing, or, 2) a search which cannot be conducted, or can be conducted very imperfectly, because of vocabulary inadequacies. Consequently, no adequate entry vocabulary has been developed.

Indexing omissions are caused by the fact that no appropriate terms are available and indexing inconsistencies also occur. This leads to the failure of certain searches that should be well within the capabilities of the system. Moreover, since searchers do not automatically inform the MeSH group of such topics, upon which they find it difficult to conduct an adequate search, these problems are perpetuated in the system.

(1) Although subheadings were apparently introduced primarily to facilitate effective use of the published bibliographies, these subheadings, as the analyses have shown, are of great potential value in reducing precision failures due to false coordinations and incorrect term relationships. The subheadings also afford an economical means of substantially increasing the specificity of the index language. The availability of free (not pre-coordinated) subheadings adds greatly to the specificity potential of the vocabulary, does not increase the size of MeSH and, by linking notions together in indexing, precludes the false coordinations that occur, for example, when the terms BLOOD PRESERVATION and PLASMA are coordinated in an attempt to express "plasma preservation."

(2) Extensive vocabulary changes tend to have a drastic effect on the economics of the search process. It is time-consuming to establish that to conduct a comprehensive search on the epidemiology of a particular disease, a certain set of terms must be used for the 1964 material, others for 1965, and add subheadings for the 1966 and subsequent material. A possible solution worth investigating is the use of automatic term substitution by the computer. For example, in conducting a search on "circadian rhythms" the computer program could cause the substitution of the term PERIODICITY for "circadian rhythm" to retrieve articles prior to the introduction of the specific CIRCADIAN RHYTHM.

K. Functions Compartmentalized

Lancaster feels that some of the problems relating to indexing (unit operations 1 and 2), searching (unit operation 6), and index language (unit operation 2), stem from the fact that these functions tend to be compartmentalized at NLM

The Index Section, the Search Section, and the MeSH Group, although they may meet periodically to discuss various problems, are self-contained units that appear to operate largely independently. The prime goal of indexing is, presumably, to describe documents in such a way that they may later be retrieved in response to requests for which they are likely to contain relevant data. However, the great majority of the indexers do not prepare searching strategies, and no mechanism exists to keep the indexers informed on the types of requests being put to the retrospective search system. Likewise, the analyses have shown that searchers are not fully aware of indexing protocols. A search on "premature rupture of the fetal membranes" was conducted on RUPTURE and RUPTURE, SPONTANEOUS, whereas most of the relevant literature is indexed under

FETAL MEMBRANES and LABOR COMPLICATIONS

or

PREGNANCY COMPLICATIONS

and the indexers claim that the "rupture" terms are inappropriate to this search since they refer to traumatic rupture, whereas "premature rupture" is a normal physiological process. Again, indexers appear to be using the term ABNORMALITIES for "process," but the analyst who prepared the formulation for a search does not seem to know this. Likewise, kidney and kidney disease terms were coordinated with DIABETES INSIPIDUS to express "nephrogenic diabetes," but it has not been the indexing policy to use kidney terms in this case.

L. Lack of Cooperation Between Indexing and Searching (Unit operations 1, 2, and 6)

From the observations of Lancaster, during the conduct of the test, it appears that the relationship between indexing and searching is not one of full cooperation towards a mutual goal. The indexers claim that searchers are "not using the correct terms"; the counter-claim of searchers is that they must "compensate for indexing inadequacies." The further separation of Medical Subject Headings (MeSH) from both the indexing and the searching functions, which has resulted in the failure to base vocabulary development on inputs from indexers and searchers, is felt to be no more healthy than the divorce of indexing and searching.[12]

The tendency towards compartmentalization of indexing, searching and MeSH development has been noted before. This is evident in the following: request analysis and search failure analysis have not been major inputs to MEDLARS vocabulary control; the entry vocabulary, which should be an integral part of the MEDLARS index language and an essential tool of both indexers and searchers, has been neglected; searchers are not completely aware of indexing policies and conventions; the average indexer has little idea, as far as the demand search function is concerned, of what he is indexing for, i.e., the types of requests that are made of the system. Lancaster recommends a close integration between the functions of indexing, searching, and vocabulary control.[13]

---

[12] Ibid., p. 99.

[13] Ibid., p. 200.

M. Conclusions of Lancaster Study

"A single evaluation study, however comprehensive, cannot be expected to discover more than a very small fraction of the specific inadequacies of the system . . . Such specific inadequacies can only be discovered through continuous monitoring of the MEDLARS operations.

"[It is] . . . recommended that the library, having concluded a large-scale study of the MEDLARS performance, should now investigate the feasibility of implementing procedures for the 'continuous quality control' of MEDLARS operation . . . [It is] recognized that continuous quality control is likely to be much more difficult to implement than a one-time evaluation. Nevertheless . . . [it is felt] that continuous system monitoring is ultimately essential to the success of any large retrieval system."[14]

---

[14] Ibid., p. 201.

VIII. FACTORS AFFECTING MEDLARS PERFORMANCE

In the foregoing study of the Lancaster Evaluation, we have restricted ourselves in the area of the Subject Indexing Function of MEDLARS, relating to the unit operations of analysis, vocabulary control, and search strategy formulation, because this is the function which has been selected for the purpose of this dissertation. The causes for this function's component failures or inadequate performance as identified by Lancaster have been brought into relief.

However, it is important to look into Lancaster's enumeration of the factors that adversely affected the overall performance of MEDLARS.[1] The test results have shown that the system is operating, on the average, at about 58% recall and 50% precision. On the average, it retrieves about 65% of the major value literature in its base at 50% precision. By extrapolation from tests, Lancaster hypothesizes a generalized MEDLARS performance curve as shown in Figure 17.[2]

The fact that, on the average, MEDLARS is operating at 58% recall and 50% precision indicates that, consciously or unconsciously, the MEDLARS searchers choose to operate in this general area. It would be possible for MEDLARS to operate at a different performance point on the recall/precision curve of Figure 17 . The searchers were on their own in making this choice. "In actual fact," Lancaster says, "we know very little about the recall and precision requirements and tolerances of MEDLARS users. This has been a much neglected factor in the design of all information retrieval systems."[3] Recall needs and precision tolerance will

---

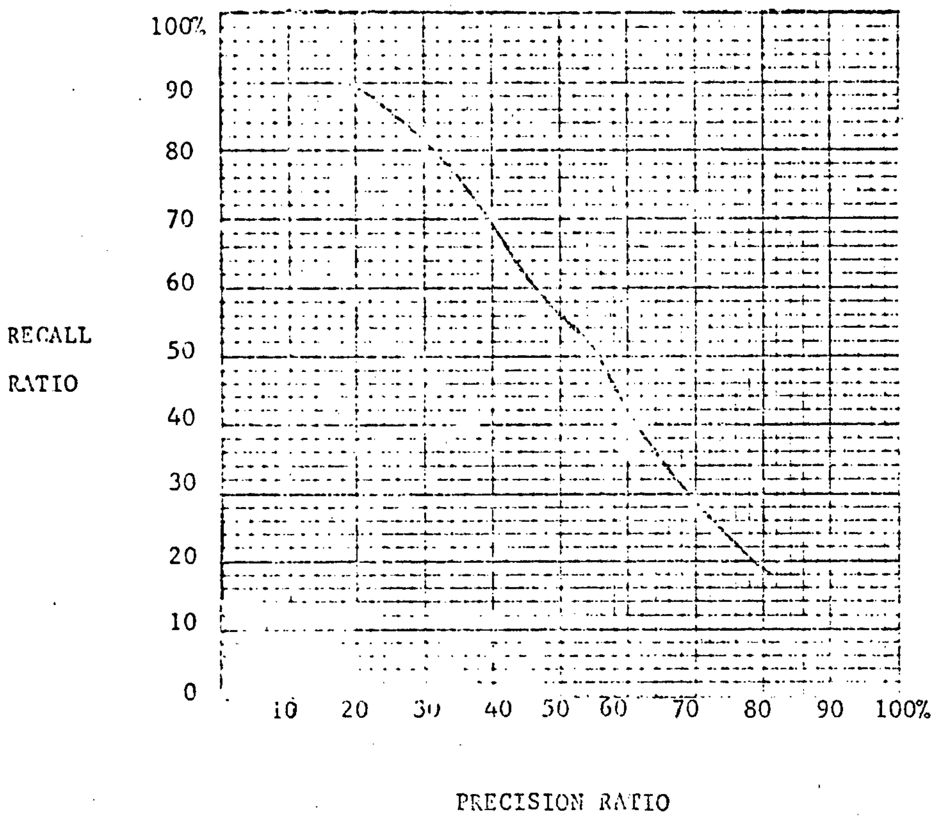
[1]*Ibid.*, pp. 185-202.

[2]*Ibid.*, p. 187

[3]*Ibid.*, p. 188

Figure 17

GENERALIZED MEDLARS PERFORMANCE CURVE



vary considerably from requester to requester, depending upon the purpose of the request, and consequently the system should be able to react to each individual request accordingly. It is important, therefore, that the MEDLARS demand search request form be so designed that it establishes for each request the recall requirements and precision tolerances of the requester, thus allowing the searcher to prepare a strategy geared as required to high recall, high precision, or some compromise point in between.

A. User-System Interaction

The greatest potential for improvement in MEDLARS exists at the interface between the user and the system. Twenty-five per cent of the MEDLARS recall failures and 16.6% of the precision failures are attributed, at least in part, to defective interaction. It is obviously crucial to the success of a MEDLARS search that a request should accurately reflect the actual information need of the requester.

B. The MEDLARS Index Language

A thorough reappraisal of the methods presently used to update MeSH is needed. There should be a shift in emphasis away from the external advisory committee on terminology and towards the continued analysis of the terminological requirements of MEDLARS users as reflected in the demands placed upon the system. As part of the quality control procedures, the MeSH group, in cooperation with the search section, should undertake the continuous analysis of MEDLARS search requests with a view to identifying areas of weakness in MeSH and legitimate requirements that cannot presently be satisfied because of inadequate terminology.

Lancaster argues that the MEDLARS entry vocabulary be regarded as an integral part of the index language of the system of no less importance than MeSH itself. Surveillance of the entry vocabulary should be the joint responsibility of the MeSH group and the Index Section. The entry vocabulary should be continuously updated and should be as easily accessible as MeSH, by the indexers and searchers alike.

Lancaster feels that there should be more use of subheadings and supports the present trend away from pre-coordinated terms (e.g., BLOOD PRESERVATION) in MeSH, to the more flexible approach of optional pre-coordination, at the time of indexing by means of subheadings. The search analyses have revealed that improved check-tags are needed to distinguish between articles such as experimental and clinical.

C. The MEDLARS Searching Strategies

The repeated reconstruction, and copying down, of strategies for notions that tend to recur frequently in MEDLARS searches is considered to be most uneconomical. Vocabulary changes have increased the complexity of searching through the different periods of MEDLARS data base. Automatic term replacement by the computer has been suggested as a possible way out of this problem.

The individual searcher makes a fairly arbitrary decision as to what type of strategy to adopt: one to aim for high recall ratio or one to aim for high precision ratio. The redesigned search request form, reflecting recall/precision requirements and tolerances of users, should enable the searchers to prepare search formulations matched to the requirements and tolerances.

Expenditure of time and effort by search analysts on citation printouts to make relevance predictions that will closely replicate the value judgments of the requester himself on seeing the actual articles is not justified. Strangely enough, knowing that relevance predictions by analysts do not closely coincide with the value judgments of the requesters, the amount of search reformulations that appears to take place at NLM is surprising.

D. The MEDLARS Indexing

The decision as to what level of exhaustivity to adopt is a difficult problem relating to indexing policy. Lancaster evaluation data in this regard have been thoroughly discussed in the previous chapter, and are recapitulated below:[4]

(1) Only a very much higher level of exhaustivity of indexing would allow the retrieval of a significant number of the relevant "depth" articles that are missed because they are not indexed with sufficient terms. Thirteen of these articles (originally indexed at an average of 7.2 terms) were re-indexed (at an average of 9.1 terms), but only two (15.4%) would have been retrieved on the re-indexing. In the other articles, the "relevant" section is very minor and would probably only be covered if the average term assignment was raised dramatically (say to 25-30 terms).

(2) On the other hand, approximately 30-40% of all the relevant "non-depth" articles that are presently missed by MEDLARS searches would be likely to be retrieved if these articles were indexed with an average number of terms comparable to the "depth" average.

Lancaster also has reason to believe that, all other things being equal, the MEDLARS recall ratio for depth articles is 70%, whereas the recall ratio is only 54% for non-depth articles.

---

[4] *Ibid.*, pp. 198-99.

Moreover, as previously noted:

(1) The division by journal into "depth" and "non-depth" creates indexing anomalies. Some of the "non-depth" articles are clearly under-indexed while some of the "depth" articles are clearly over-indexed.

(2) Because of term limitations, some of the non-depth articles are indexed in such general terms that it is difficult to visualize a single search in which they would be retrieved and judged of value. In other words, these citations are merely occupying space on the citation file.

To recapitulate, we can say: a substantial number of recall failures occur due to lack of exhaustivity of indexing; a marginal increase in the average number of terms assigned to "depth" articles is unlikely to result in any significant recall improvement while a major increase is unjustified on economic grounds; raising the present "non-depth" level to the present "depth" level is likely to result in a 30-40% improvement in retrieval of relevant articles from non-depth journals; the present division of journals into "depth" and "non-depth" has led to indexing anomalies and to the situation in which non-depth articles occupy 45% of the file, but account for only 25% of the retrievals; some of the non-depth articles are never likely to be retrieved and judged of value because they are indexed much too generally.

On the basis of the above, Lancaster recommends that the present distinction between "depth" journals and "non-depth" journals be abandoned. This does not mean that all articles from the present non-depth journals should be assigned an average of ten index terms. Rather, it means that each article should be treated on its own merit and sufficient terms should be assigned to index the extension and intension of its content.

Lancaster sees no justification for an overall increase in indexing exhaustivity at the present time.

Although few indexing errors (in the sense of incorrect term assignment) were discovered in the evaluation, a significant number of indexer omissions were encountered. Indexer omissions accounted for approximately 10% of all the recall failures. However, some of these indexer omissions appear to be largely due to lack of specific terms in the vocabulary. If no specific term is available for a concept, either in MeSH or in the entry vocabulary, an indexer is quite likely to omit it entirely (rather than trying to cover the topic in a more general way). Lancaster believes that indexer omissions will be substantially reduced as the entry vocabulary is improved.

Moreover, a very small spot-check (reported earlier) suggests that perhaps 25% of the failures attributed to indexer omission might not be the fault of the indexers, but might be due to the deletion of a term after the indexer has assigned it. This is discussed further below.

E. Computer Processing

Computer processing was not a major cause of retrieval failures in the study. However, there has been one situation where it appears that a term was deleted by some faulty file maintenance procedure. The system must have the ability to check against any deletion of this sort, and have adequate file protection mechanism.

F. The Relationship Between Indexing, Searching and MeSH

The tendency towards compartmentalization of indexing, searching and MeSH has been noted in the previous chapter. A close integration between the functions of indexing, searching, and vocabulary control is needed.

G. Use of Foreign Language Material in MEDLARS

It has been noted that while foreign language articles consume approximately 45% of MEDLARS input costs, they contribute no more than 16% of the total demand search usage. This is a major policy problem. It may be useless retrieving foreign language citations without backing them up by providing adequate translation facilities.

H. Search Printout as a Content Indicator

It has been found that titles and tracings are frequently inadequate in indicating the content of articles. In the light of this, the requirement for inclusion of abstracts in the data base is indicated.

To recall the conclusions of Lancaster, "A single evaluation study, however comprehensive, cannot be expected to discover more than a very small fraction of the specific inadequacies of the system . . . Such specific inadequacies can only be discovered through continuous monitoring of the MEDLARS operations."

This is why Lancaster recommended that the library, having concluded a large-scale study of the MEDLARS performance, should now investigate the feasibility of implementing procedures for the "continuous quality control" of MEDLARS operation. Lancaster recognized that continuous quality control was likely to be much more difficult to implement than a one-time evaluation. Nevertheless, he felt that continuous system monitoring is ultimately essential to the success of any large retrieval system.[5]

---

[5] Ibid., p. 201.

I. Relevance of PERT/CPM

In Chapter II we have seen that PERT/CPM is a networking technique with time estimation and cost computation capabilities. It can identify the network nodes in a context of precedence and dependency relationships, and determine the critical path through the network. This is the time-cost based, graphical representation of a system.

In scheduling we have found the technique of handling the "input---> processing---> output" operations of a system component or network node. A system component receives input from another component belonging to the system, operates on the input, and produces an output which becomes the input of another component. This is a basic function performed by a basic functional unit--the network node. The process is repeated until the final system product or service is produced.

We have studied the characteristics of information systems and indicated their isomorphism with systems in general, and as such, the possibility of the use of PERT/CPM in the development of an information system design methodology. We then studied MEDLARS, a large-scale computer-based information system, to identify the factors that caused MEDLARS to perform its design functions in a less than optimum manner.

Now we turn to see what actions MEDLARS has taken to implement procedures for the "continuous quality control" as recommended by the Lancaster Evaluation study.

J. A Small Staff in the B.S.D. is Not the Answer

From the recent reports emanating from MEDLARS, it does not appear that the MEDLARS management is contemplating control of the system at the basic functional unit level. Under the heading "Quality Control" The National Library of Medicine Annual Report for the Fiscal Year 1968 writes, "In January 1968, Evaluation of the MEDLARS Demand Search Service, by F. W. Lancaster, Deputy Chief of the Bibliographic Services Division (B.S.D.) was published by the Library. This evaluation, based upon a thorough study of 300 demand searches, is a source of much useful information concerning the strengths and weaknesses of MEDLARS as a bibliographic citation retrieval system during 1966 and early 1967, when the study was performed. MEDLARS is a dynamic system in every respect. The staff involved in all phases has expanded greatly. The vocabulary and many other aspects of the system have been undergoing rapid change. In order to access current system performance, and to identify factors tending to produce irrelevance or incompleteness in MEDLARS products, an ongoing evaluation must be maintained. During fiscal year 1968, plans were developed for a small staff, in the Office of the Chief, Bibliographic Services Division, to monitor MEDLARS quality, including the quality and the consistency of indexing, as well as the characteristics of the searches and bibliographies produced. This staff is expected to concentrate its efforts on providing information as a basis for inaugurating improvements in system procedures and practices. This group will also do the preparatory work that is required to allow NLM to derive the greatest advantage from deliberations of the Committee on Selection of Literature for MEDLARS, the advisory group concerned with quality of the literature indexed for MEDLARS."[6]

This is not incorporating control at the "cellular" level of the system "physiology"; this is establishing an office of control "to monitor MEDLARS quality." This staff will have no direct involvement in

---

[6] The National Library of Medicine Annual Report for the Fiscal Year 1968 (Washington, D.C.: Government Printing Office, 1969), pp. 31-32.

the continuous operations of the basic functional units and its actions will have to wait until something that warrants control action surfaces, overcoming the "gravitational pull" of the hierarchy. In the quotation above, MEDLARS has been called "a dynamic system in every respect." In a dynamic system, errors compound faster, and to maintain the dynamic equilibrium of an open system, continuous control at the basic functional unit level appears to be sine qua non.

K. Conclusions from the Lancaster Study

Although the original MEDLARS philosophy was to perform all indexing centrally with NLM staff, the massive volume of work to be done, coupled with rapidly increasing backlogs, caused library management to reconsider this policy and begin to use outside contractors for some of the indexing work. It appears that application of sequencing and queuing techniques would have predicted the backlog by indicating the rate of growth of the queue and the inadequacy of the service points, and that a control mechanism could be developed which would alert the responsible component of the system (here the management) to take corrective measures before the development of the backlog.

It seems that a control mechanism incorporated in the basic functional unit of the system can continuously monitor the unit's performance and keep correcting the unit's operations against a "pre-set value" so that the situations like the vocabulary inadequacy, as pointed out by Lancaster, may be corrected in "real-time" instead of waiting for the accumulation of error data for a considerable period of time and then taking the necessary corrective measure when, maybe, it is already too late. The purpose of the incorporated control mechanism is to make the system behave like an adaptive system.

Lancaster was given ten principal objectives for the Evaluation study with regard to Index Language and Indexing, such as--"Are there significant variations in inter-indexer performance?" and so forth. These questions could have been posed in "real-time" and corrective measures could have been taken if there were control mechanisms at the basic functional unit level.

The case of "tetrodotoxin" and recording decision in the MEDLARS entry vocabulary also suggest the possibility of real-time action.

A significant number of cases of indexer omissions can be attributed to the fact that no MeSH term exists for the missed notion, and there is nothing in the entry vocabulary to say how the topic is to be indexed. As a result, the indexer either omits the topic entirely or indexes it much too generically. This is a case which could have been corrected by control action if control mechanism were available at this level.

Lancaster was surprised to discover that very little use was made of weighting as a retrieval device although MEDLARS has a built-in <u>term weighting system</u>. This is a clear case of lack of use of an available capability and lends itself to control action, provided the mechanism is there at the level where the function is taking place.

Within the Evaluation program, requests have been systematically analyzed from the point of view of the capability of the vocabulary to cope with them, but this is not done as part of the regular operations of the system. Although a form is available to record suggestions of indexers and searchers, very little use appears to be made of this. There are no routine established procedures whereby indexers and searchers are required to notify the MeSH group of vocabulary inadequacy. Indexing omissions are caused by the fact that no appropriate terms are available, and since searchers do not automatically inform the MeSH group of such topics, upon which they find it difficult to conduct an adequate search, these problems are perpetuated in the system. There could be no more justification for control mechanism at the basic functional unit level than this!

Functions tend to be compartmentalized at NLM. Self-contained units appear to operate largely independently. Indexers do not prepare search strategies, and no mechanism exists to keep the indexers informed on the types of requests being put to the retrospective search system. Likewise, the analyses have shown that the searchers are not fully aware of indexing protocols.

These are not subjective problems. They are perfectly tractable and may be subjected to control action. But they lingered in the system because there was no control mechanism at the basic functional unit level to alert any component to take corrective measures in "real-time."

The foregoing analyses based on the Lancaster study indicate that large-scale computer-based information systems cannot function properly without CONTROL at the basic functional unit level. This is why Lancaster recommends a feasibility study for implementing procedures for the "continuous quality control" of MEDLARS operations.[7]

It is argued that control is the essence of all successful organization and that the control mechanism resides in the basic functional units of the system, serving as coordinator, regulater, stabilizer, or governor. A system is obtained by networking the basic functional units, which integrate into the desired system.

There is obviously a <u>need</u> for an information system design methodology which can handle the problem of incorporating CONTROL in the basic functional units of the system components, which are ultimately networked into the desired system.

[7]Lancaster, Evaluation of the MEDLARS Demand Search Service, p. 201.

The "Activities" of a PERT network are analogous to the basic functional units of an organization or system and the "Events" of a PERT network can be compared with outputs of the units. The control mechanism will take a fraction of an output "error," if any, and utilize that as a stimulus or lever to activate another component for compensatory action to stabilize the system behavior.

Adequate records of the network activities must be maintained and used as input to the control system. The records will be generated by the process itself, as the network activity will receive input, operate on it, and produce an output. The output "error," if any, will be, in reality, a record of the error and will be used as input to the control system.

As we have seen, PERT/CPM forces us to set the system components in a precedence and dependency relationship context. Thus the activity flow of the system is controlled. The scheduling that takes place inside a component controls the internal activity, and since this is a basic function of the system performed by a basic functional unit, control is established at the basic functional unit level. Since these basic functional units are interdependent, activity of one is affected by the activity of the other in a predefined manner (i.e., defined by the dependency), and malfunction in one will immediately trigger the control mechanism of its relational constituents.

Normally, once a system has been designed and implemented, PERT/CPM is dropped. But as we are using it in this dissertation, it is a graphical representation of the physical system existing and moving in parallel at all times, during the design, implementation, operation, and evaluation of the system keeping it always in sharp definition.

IX. A CONTINUOUS MONITORING DESIGN METHODOLOGY

An information system is expected to perform its design functions. This is important for the information system evaluator to remember. The users of an information system may have a variety of information needs. But the system may not have been designed to meet all of them. A system may be over-designed or under-designed. An example of over-design will be the publication subsystem for recurring bibliographies in MEDLARS. The original estimate of 50 recurring bibliographies was too high, and only nine were in production on January 1, 1968. Nevertheless, the subsystem was designed to cope with 50 recurring bibliographies. On the other hand, a system may be required to handle a thousand profiles when it has been designed for an anticipated load of only one hundred.

To be fair, a system must be evaluated on its own terms. But when that is done, and any system or component failure is detected, then the cause of the failure must be attributable to some design deficiency. This is quite normal and expected. No one can design a complex, computer-based large-scale information system, anticipating all possible exigencies so that nothing will ever go wrong. On the contrary, as we all know, if anything can go wrong, the chances are that it will go wrong at the most critical time.

But deficiencies may be corrected if they can be detected. So we evaluate systems at intervals. A medical analog of this would be an occasional physical checkup for possible diagnosis, therapy, and prognosis. The Lancaster evaluation of MEDLARS typically exemplifies this approach. The system ran for a while then Lancaster evaluated it. He came up with his conclusions and recommendations.

This would be just like any other evaluations unless he had come up with a very interesting conclusion which had nothing to do with the system and its operation. The conclusion is about the process of evaluation itself.

To recall, he concluded that "A single evaluation study, however comprehensive, cannot be expected to discover more than a very small fraction of the specific inadequacies of the system . . . Such specific inadequacies can only be discovered through continuous monitoring of the MEDLARS operations"; and recommended that the "Library (NLM) . . . should now investigate the feasibility of implementing procedures for the 'continuous quality control' of MEDLARS operation." He recognized the difficulty of implementing continuous quality control, but, nevertheless, felt that "continuous system monitoring is ultimately essential to the success of any large retrieval system."

Lancaster's admonition can hardly be overemphasized. Evaluation studies like the one MEDLARS had, can only be of historical or archival interest. Information systems are open and dynamic. Both the system components and their interrelations change with time, making most of the evaluation findings contextually irrelevant. "As Calvin Mooers pointed out in a meeting of the MEDLARS Evaluation Advisory Committee, whatever changes might be made in the future, there are some half-million citations in MEDLARS and it would be some years before a change in, for instance, present indexing policy could be expected to have any major effect on the overall performance."[1]

Saul Herner supports Lancaster when he reflects about evaluation and maintains that, "If it is done effectively--if it is thought of as a

---

[1] *Ibid.*, p. 128.

matter of quality control--it is a continuing, and never a one-shot process. In a dynamic situation . . . requirements change, methodologies and technologies change; the best way of meeting a requirement now may become comparatively inefficient later. People operating systems change, and machines or mechanisms get old or obsolete. And so we can never afford to be sanguine about systems. We have to incorporate continuous and rigorous quality control procedures into their operations. That is the only way we can be sure we are doing the job we set out to do: to meet the existing information needs of our audience."[2]

So the Lancaster evaluation has given us diagnostics on systems problems of an operational information system and emphasized the need for "real-time" control of systems. In our discussion of the "Problems of Information Systems" we asked ourselves a couple of questions, namely, 1) Is it possible to develop design requirements from the diagnostics generated by the system operating experience and create design algorithms which will force the designer to go through the process of problem solving at the point of their logical occurrence on the drawing board? and 2) Is it possible to develop a design methodology which will also provide mechanisms for trouble-shooting as they will occur at the basic functional unit level? We refrained from trying to answer those questions because at that point we did not know enough of the problems of information systems.

Now we know about the problems of information systems in general, and the evaluated operating experience of an on-going information system in particular. We also know that there are techniques available with which we can isolate the basic functional units of a system and set them

---

[2] Saul Herner, System Design, Evaluation and Costing---in Plain English, Contract No. AF49 (638) - 1424, Project No. 9769-01 (Washington, D.C.: Herner and Company, 1969), p. 14.

in a time/cost, precedence and dependency relationship network. Unless they are redefined, the basic functional units remain the same. However, the network configuration representing their interrelationships may change in real-time depending upon the exigencies of operating experience.

We have also seen that techniques are available for monitoring the internal activity of the basic functional unit as it processes the input received from the preceding unit and produces an output for the successor unit, by the application of assignment and sequencing algorithms.

Let us see what all this is doing for the designer. These techniques seem to give the designer the capability to control the time, control the cost, manipulate interrelationships, and to control the internal processing of an activity, and all this in real time, because these techniques cannot be used in any other way than in real time.

But we need to test this. In other words, we need to test the hypothesis that PERT/CPM methodology or some modified version thereof can be developed into an Information System Design Methodology.

To do this, first of all we will have to redefine a PERT activity and introduce some modifications to suit our purpose. Then we will isolate the activities of a hypothetical information system and network them into the desired system structure. This initial blanket network will be called "umbrella net." This network will provide a panoramic view of the total system from the initiation stage to the final disposal stage.

The subject indexing function of MEDLARS has been selected for this dissertation. It has been stated before that proper operation of this function is probably the most important single factor governing the performance of an information retrieval system. It would not have mattered, however, if any other subset of the system had been selected. This subject indexing function of MEDLARS will be identified with its counterpart in in the umbrella net, and a PERT network of this function will be created based on MEDLARS system description and data flow charts.

Eventually we will focus on only one activity by the application of a family networking technique and go through some micromanipulation with reference to the "modified PERT activity."

After this we will present the PERT Computational program, CPM Computational procedure, and the Scheduling Model (presented in two parts as the Assignment Model and Sequencing Model), in that order.

The PERT Computational program will compute the time estimates for the network activities and identify the critical path through the network, thus providing control over time. The CPM computational procedure will help in making the decisions between the time/cost alternatives and hence provide control over cost.

The Assignment Model will help in the optimization of assigning jobs to capabilities, and the Sequencing Model will optimize the handling of jobs which need different treatment on different equipment in different order or sequence. These two Models together will provide control over the internal input processing and output generation of activities. With adequate record keeping, some redundancy, and the redefinition of PERT activity, it will be seen that PERT/CPM/scheduling methodology can be developed into an Information System Design Methodology.

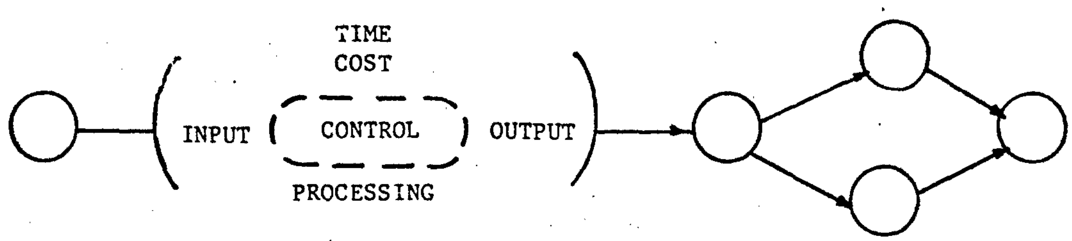
A. The Demonstration

PERT Activity Modified and Redefined. According to the design of the "experiment" as laid out in the previous section, we now have to redefine a PERT Activity and introduce some modifications to suit our purpose. Then we will isolate the activities of a hypothetical information system, its design, implementation, operation, and evaluation, and network them into the desired system.

A PERT activity is a time consuming operation which receives an input, operates on it, and then produces an output, which is an event, and which becomes the input for the next logical activity. The only exceptions to this are the lead and end events.

As we have seen before in the review chapter on PERT/CPM, normally a PERT activity would have three time estimates--optimistic, most likely, and pessimistic. A PERT event is considered as the output of the PERT activity.

Now let us see in what respect the PERT activity should differ from the normal to serve as the basic functional unit of the system. We know that PERT provides time estimates, and CPM computes cost. But they can provide these estimates having received the input, processing this information, and producing an output as shown in Figure 18 below.



INPUT-PROCESSING-OUTPUT WITH CONTROL

Figure 18

As the PERT/CPM technique will tell us how long the activity is going to take and how much it is going to cost, it should also, at the same time, operate on the input and provide enough information for assigning and sequencing the input to produce the necessary output, and to determine how it is going to deliver the output to the next logical basic functional unit.

For continuous control, the control mechanism must reside in the basic functional unit. In the previous section we have mentioned how the PERT computational program will compute the time estimates for the basic functional units (the network activities) and identify the critical path through the network, thus providing control over time. The CPM computational procedure will help make the decisions between the time/cost alternatives and hence provide control over cost. The Assignment Model will help in the optimization of assigning jobs to match available resources and the Sequencing Model will guide the tasks (or jobs) for processing in the sequence which matches the requirements of each specific task. These two Models together will provide control over the internal processing of input and produce an output of activities. Thus the "modified" PERT activity will look like the following Figure 19.

Figure 19 redefines and modifies the PERT activity and will imply all this whenever the word activity is used, unless otherwise specified or the context makes the meaning obvious.

Processing is the actual work that is accomplished in an Activity. The work is divided into tasks or jobs and routed through the men and machines, matching the task requirements and men and machine capabilities in some order where applicable. These are Assignment and Sequencing problems or, in other words, optimum allocation of resources problems that can be handled by the application of Operation Research techniques such as Assignment and Sequencing. An Assignment and a Sequencing model have been adapted in this dissertation.

Structure, properties, rate, and frequency are the attributes of both input and output. Structure and properties can be determined by physical study and analysis of both input and output. The study should

THE MODIFIED PERT ACTIVITY

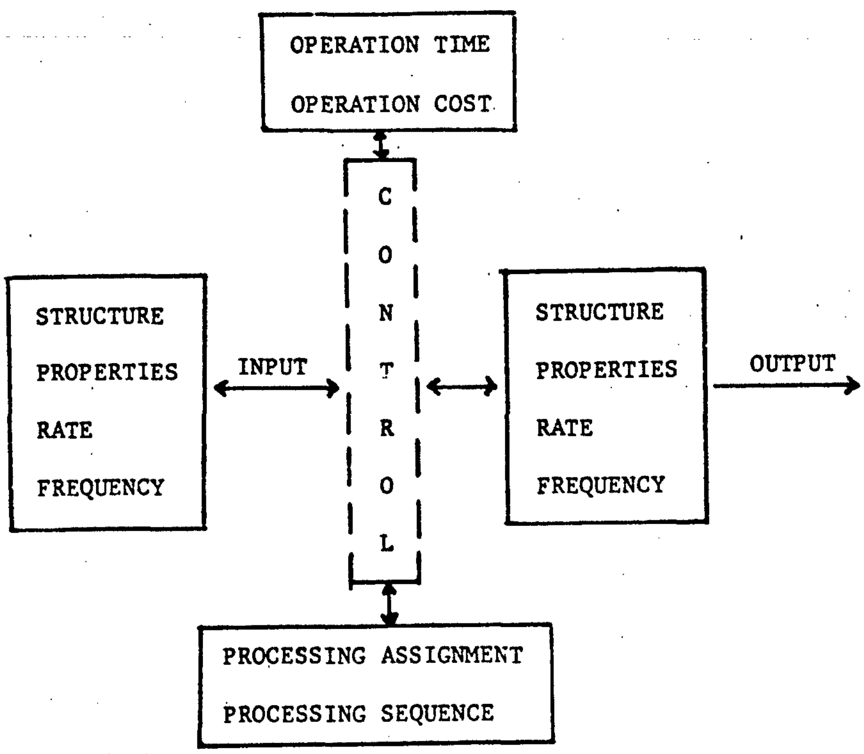


Figure 19

answer questions like, is it erasable? Is it easily perishable? What is its volume, unit size, etc.?

Rate and frequency for both input and output can be studied by the application of Queuing Theory, Markov Process, Poisson distribution, and similar techniques.

Operation time and operation cost are the two interrelated parameters of an Activity and can be studied by the application of economics of scale, crash-normal-and in between time/cost estimations and similar micro-economic techniques. The other techniques and methods mentioned above are available in the literature and experimental application of these will constitute important and urgently needed research in the library and information science area.

B. Networking the Activities into the Desired System

We have now redefined and modified the PERT activity. At this point, we will start networking the activities into the desired system.

The initial blanket or umbrella network that the designer will start with must provide a panoramic view of the total system from the initiation stage to the final disposal stage. It must also show the precedence and dependency relationships among the events and activities making up the network.

Each of the activities and events of the Umbrella Net will become a series or family of networks of descending generality as the design process will be unfolding. Whenever necessary, ligands may be formed by combining two (maybe more) nodes of different sub-network systems indicating their relationships. Any delivery to the system is an example of this ligand formation.

activities will be identified and isolated through systems analysis. These activities are the means to accomplish objectives. The analysis of the need will generate the system objectives. The activities are merely the means selected from amongst the available alternatives to meet the need.

Following is the itemization of the activities of the hypothetical information system, its design, implementation, operation, and evaluation, as identified for the development of the Umbrella Net.

Activity 0-1: Establish information system for the designer. "Creativity is essentially a process of making new combinations of known pieces of knowledge; a new idea is not just imagined, it is produced by synthesis, or at least by analogy with known facts."[3] The process of designing is partly creative and partly algorithmic.

Efforts of Norris (1963) in developing the morphological approach to design, of Jones (1963) in developing the logical approach, and of Latham (1965) in developing PABLA (Problem Analysis by Logical Approach), and finally of McCrory, Wilkinson and Frank (1963) in comparing scientific research methods with the steps of determining the need, analysis of the need, design conceptualization, determinations of feasibility, and final production of the system, have brought the algorithmic segment of the design process into sharper focus.[4] The creative segment of the design process will have to depend on the intuition, imagination, and ingenuity of the designer.

---

[3]J. Farradane, "Information for Design," in The Design Method, ed. by S. A. Gregory (New York: Plenum Press, 1966), p. 98.

[4]Ronald D. Watts, "The Elements of Design," in The Design Method, Ibid., pp. 85-95.

However, both these segments thrive on information. The information system that is established to serve the designer should do the following:

(1) Collect, organize, and provide on demand and/or on a current-awareness basis, information bearing on the design project;

(2) Document information generated during the design process;

(3) Keep the members of the design team informed of each others work; and

(4) Generate all the instruments of communication of the design team in collaboration with the element of the design team involved.

Activity 1-2: Schedule general systems analysis. The total systems analysis is scheduled here. Time schedule is set up subsystem by subsystem for analysis. This will set up the time-table for the entire project and will take into consideration all the constraints and deadlines. It will set the general limits and guidelines within which the systems analyses have to be performed.

Activity 2-3: Estimate budget and staff required for systems analyses. Taking the limits and guidelines established in the preceding activity, a budget for the total project will be worked out. Staff requirement will be estimated at this stage, including their category, number under each category, job descriptions, and desired skills and competences. The design team is now partially formed.

Activity 3-4: Identify system objectives. "The major objectives of an information system are to bring relevant data in usable form to the right user at the right time so that they will help in the solution of the user's problems."[5] A complete array of the desired objectives and goals for the system is set up at the stage. The hopes and ambitions of the system are crystalized and documented here as targets for achievement. The objectives may include the following:

(1) Types of products and/or services to be offered, e.g., published indices, on-line access, etc.;

(2) Format, frequency, and load, e.g., 3"x5" cards as form of search output, 24 hrs. turn around time, through-put of a 100-question batch, etc.;

(3) The nature, size, and geographical dispersion of the clientele to be served;

(4) Adaptability, compatibility, and growth potential of the system; and

(5) Perspective objectives projected with an awareness of technology forecast, e.g., video-telephone access to data sets.

Activity 4-5: Select the means to attain the objectives. There is no value in having utopian objectives unattainable by the application of the current state-of-the-art. At the current point of time, there has to be a one-to-one relationship between the objectives and the means of attainment. At this stage, the hardware-software, man-machine configuration for attainment of the objectives is established. The intellectual means of attainment of the stipulated system objectives might include:

(1) Thesauri or other instruments of terminology control;

(2) Various look-up tables for performing transformations, error checks or standardization of data;

---

[5] J. Jaffe, "The System Design Phase," in Developing Computer-Based Information Systems, by Perry E. Rosove, op. cit., p. 94.

(3) Design of forms such as input forms, report forms, evaluation questionnaires, forms for recording search strategy, etc.;

(4) Intellectual manpower for systems operation; and

(5) Programming manpower for producing the software for the system.

The physical tools might include the following:

(1) Hardware for input;

(2) Hardware for output;

(3) Satellite and buffer hardwares;

(4) Main computing facilities;

(5) Data and image transmission equipments; and

(6) Algorithms and softwares for job and system control.

Activity 5-6: Set up the schedule, budget, and staff for design of the system. This step is analogous to the activity 2-3; in fact it will augment the design team by inclusion of design staff. Only it cannot occur before the knowledge and experience gained through the previous activities.

Activity 6-7: This is the stage to finalize systems specifications. This is the communication generated by the designer in response to the original communication of the need and released into the environment in the form of a set of prescriptions for the embodiment of the design. This step is not complete without the completion of activities 6-8 and 6-9, but these two activities could be parallel to activity 6-7 as shown in the Umbrella Net. (See Figure 20, pages 128-130).

Activity 6-8: Design systems administration: job description, staff, hierarchy. This is where the system's managerial and administrative requirements, both intellectual and physical, are established for a number of years after the system's initiation. Block diagrams of the system's administrative structure (organization chart) indicating hierarchy and reporting relationships will be detailed. All that is associated with staff planning, taking into account the assessed immediate and projected load and necessary budget, is determined here.

Activity 6-9: Flow-chart systems operation: input, processing, output, feedback. The system's operation is flow-charted here indicating flow direction, decision points, branch-off points, links, and interrelationships of operations. This is the graphic representation of the operations subsystem.

Activity 8-10: Schedule systems realization. Now it is time to make the system a reality. We have everything necessary for the embodiment of the system. We may build it, procure it, or adapt an existing system to meet the specifications. A schedule is set up for delivery of the components and subsystems and a target date is fixed for the system to become operational.

Activity 10-11: Systems test and adjustments if necessary. The components and subsystems, as they are delivered, must be subjected to strict scrutiny. They must pass through a quality control and reliability test procedure to guard against systems failure or less than optimum systems performance. After adjustments, if necessary, the total system is tested and okayed.

Activity 11-12: Systems initiation and operation. This is the stage when the system is launched and becomes operational. This may be called an open-ended activity and should be in progress during the life expectancy of the system. This should take into account depreciation, replacement and repair, weeding and retirement, and so forth.

Activity 12-13: Systems disposal. It is important to visualize and plan for the disposal of the system when it reaches the normal age of superannuation. There may be many possibilities falling between simple discarding and thorough rejuvenation. For an adaptive system, as we have visualized here, it may not be impossible to keep viable indefinitely through proper functioning of the feedback/control system and guarding against obsolescence by timely reparative growth, replacement, and replenishment.

These activities, as itemized above, have been networked into the Umbrella Net, indicating their precedence and dependency relationships (see Figure 20).

We now have the Umbrella Net of a hypothetical information system. We have selected MEDLARS as the object system for this study and have stated before that the subject indexing function of MEDLARS will be identified with its counterpart in the Umbrella Net, and a PERT network of this function will be created, based on MEDLARS system description and data flow charts.

Figure 21 is a PERT representation of the subject indexing function of MEDLARS. This network belongs to the activity 6-9 of the Umbrella Net, Figure 20, and lays out the different components of the function, indicating their precedence relationship.

This network has been derived by applying the "Family networking" technique (Figure 6, see p. 29) to the MEDLARS Umbrella Net as appears in the "MEDLARS System Overall Data Flow Chart," Figure 22. The first level expansion is shown in Figure 23, labelled "MEDLARS Input Subsystem Flow Chart." Here the "Indexing" block of the Umbrella Net has been expanded in the section marked "Bibliographic Services Division," separately shown in Figure 24. These flow charts and data have been taken from Austin,[6]

---

[6]Charles J. Austin, MEDLARS 1963-1967, Public Health Service Publication No. 1823 (Bethesda: National Library of Medicine, 1968), pp. 10, 14.

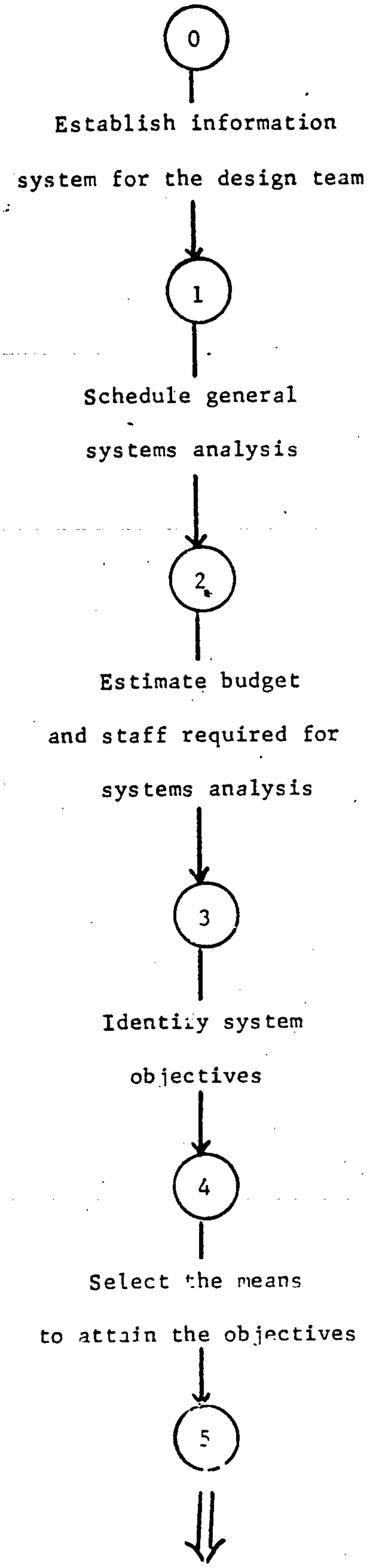


Figure 20: THE UMBRELLA NET

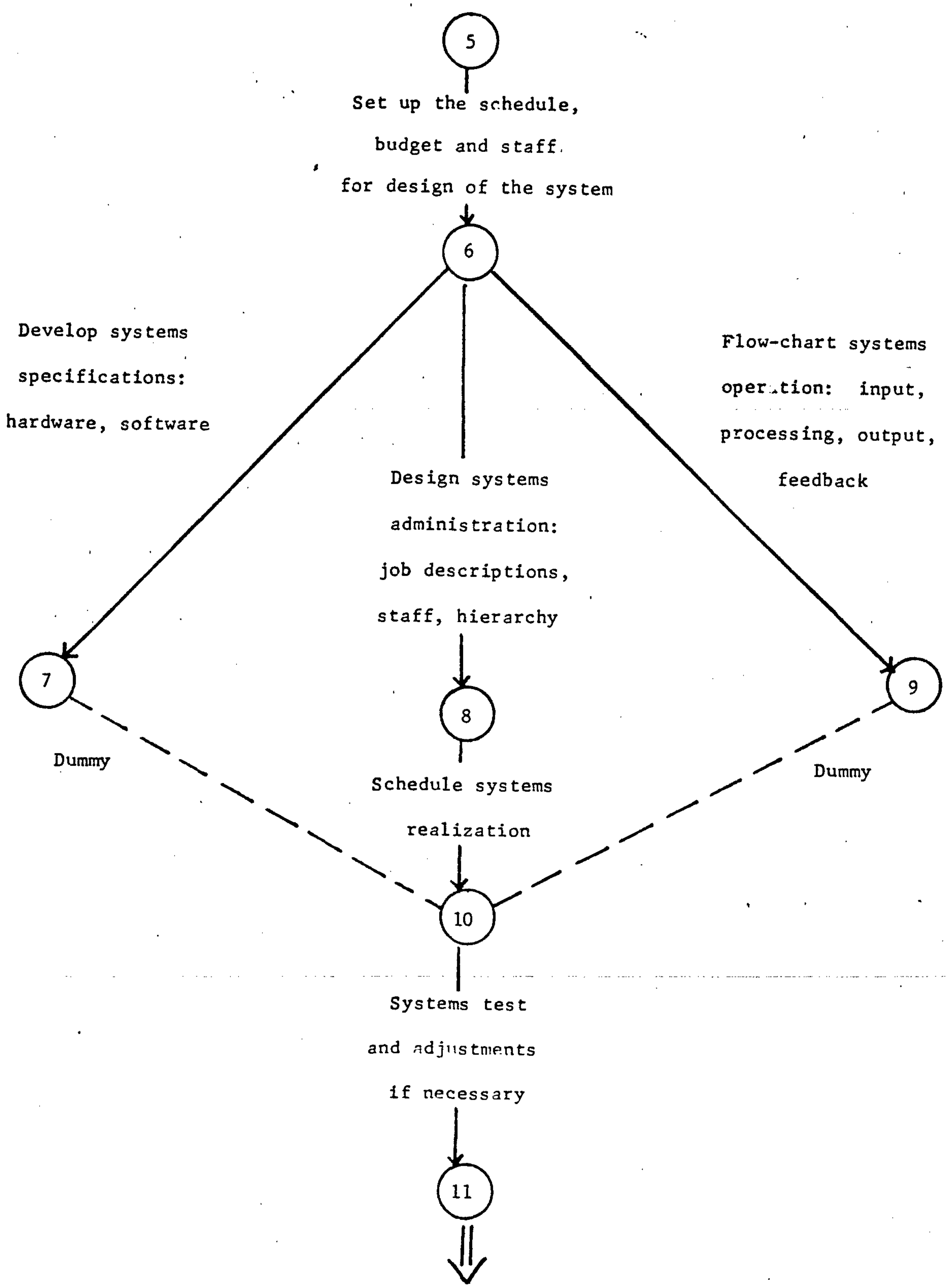


Figure 20 (Contd.)

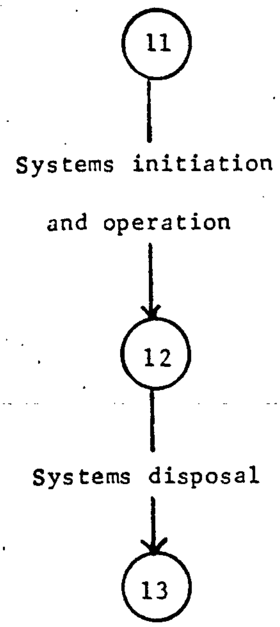
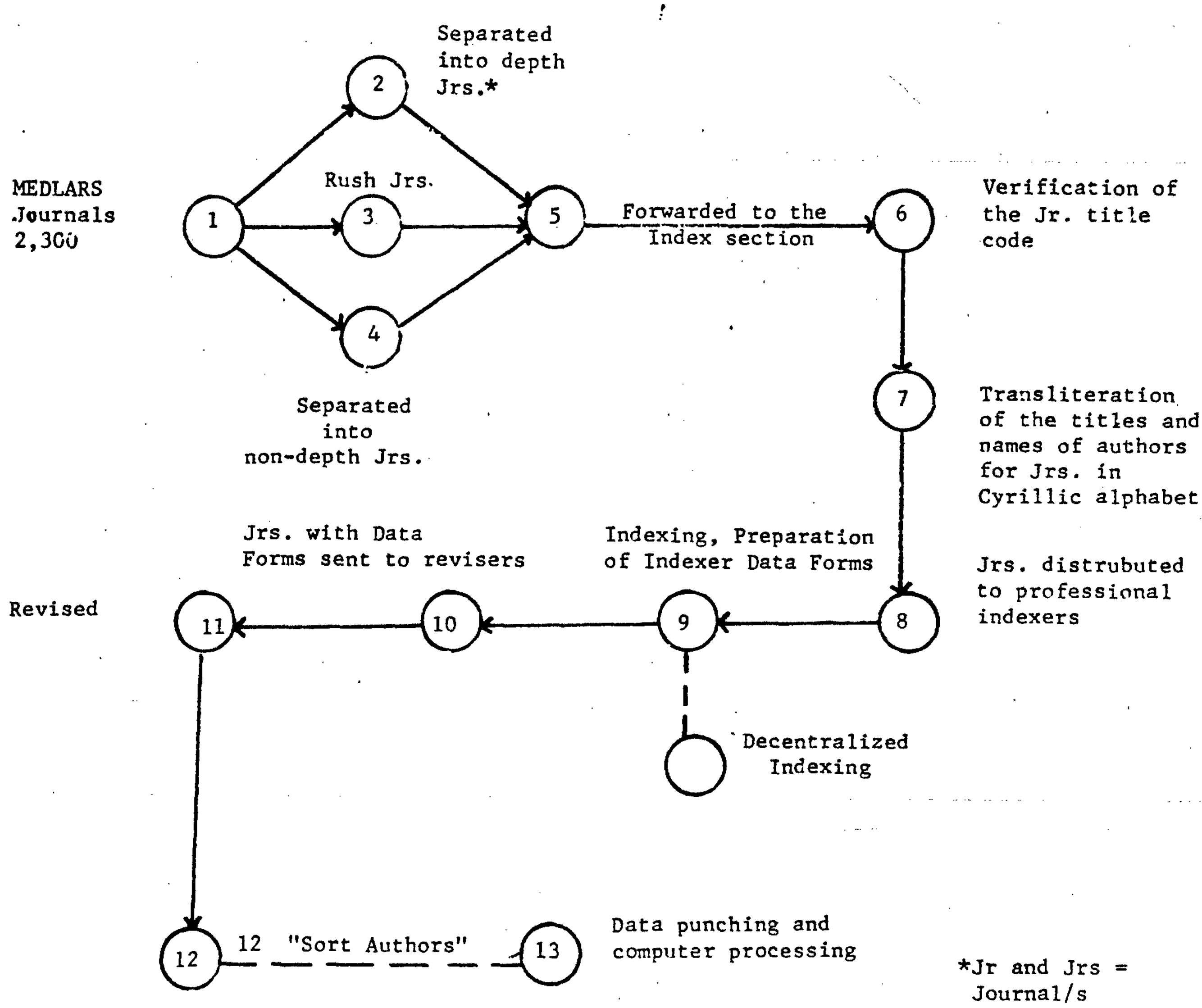


Figure 20: (Contd.)

and summarized below with reference to Figures 22-23. Where such data are not available, i.e., when an entirely new system is being designed, it will be necessary to go through the process of making decisions between possible and feasible alternatives with respect to the need.



PERT REPRESENTATION OF MEDLARS SUBJECT INDEXING FUNCTION

Figure 21

MEDLARS SYSTEM OVERALL DATA FLOW CHART

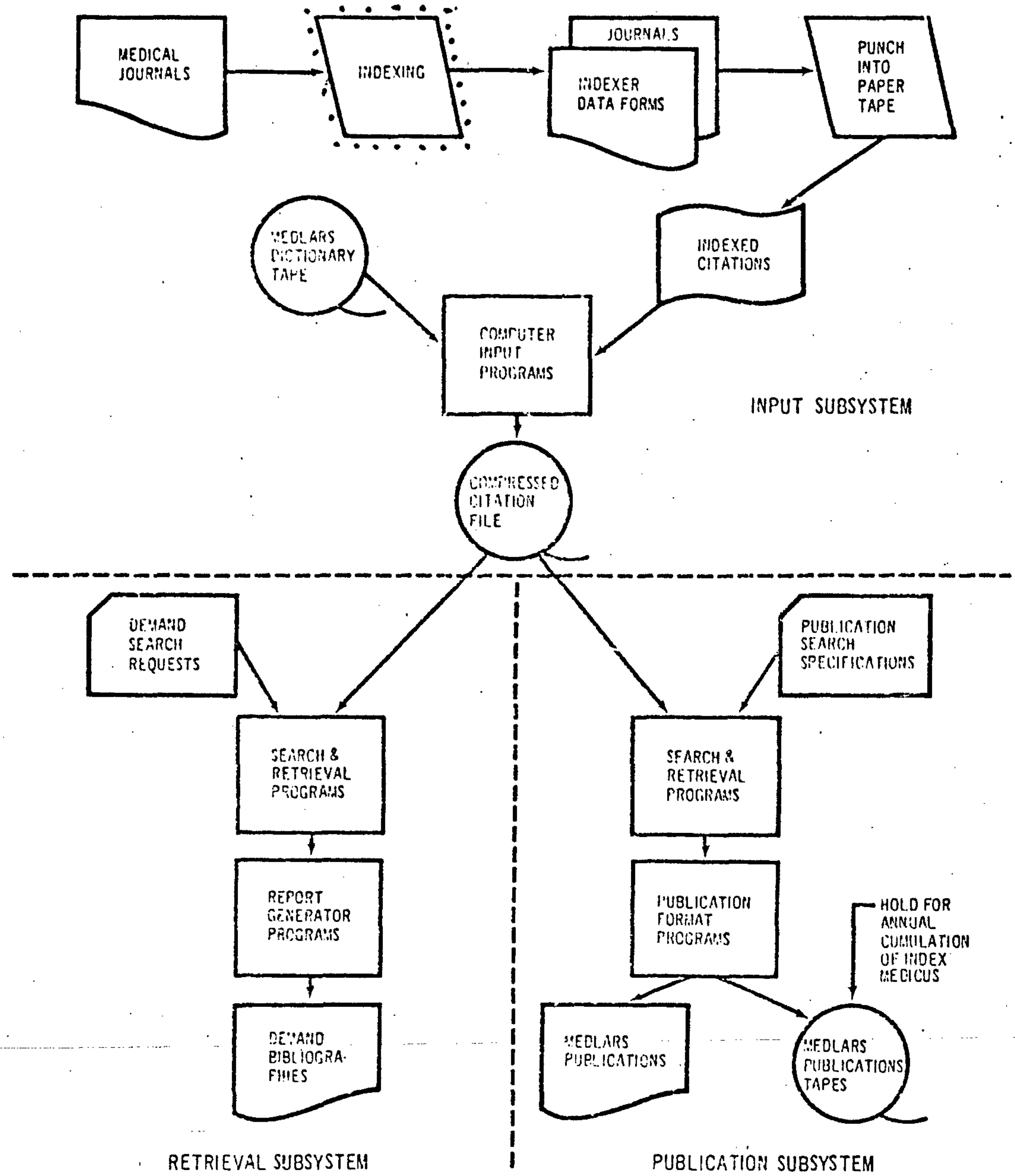


Figure 22

Figure 22 shows the overall data flow of the MEDLARS system. In the Input Subsystem (top), journals are received and indexed. Paper tapes are punched using the Indexer Data Forms. The computer input programs then use the Indexed Citations and the MEDLARS Dictionary Tape to generate the Compressed Citation File. This file is used by the Retrieval Subsystem (lower left) and the Publication Subsystem (lower right) to generate the Demand Bibliographies and MEDLARS Publications respectively. The INDEXING box is embellished because that is where our interest lies.

The Input Subsystem is the functional portion of MEDLARS concerned with selection of journal articles, indexing, conversion to machine-readable form, and input to the computer for storage on magnetic tape, as shown in Figure 23.

The National Library of Medicine currently receives between 18,000 and 19,000 different serial publications of all types. The contents of approximately 2,300 biomedical journals are indexed for input into MEDLARS.

Journals selected for input are divided into two groups based upon the scientific significance of the material published; a depth-indexing group (journals that regularly carry reports of greater significance), and a non-depth group (journals containing material of lesser significance). The depth journals are indexed in much more detail than the non-depth ones.

The MEDLARS journals are batched and forwarded from the Technical Services Division (Figure 23, top) to the Index Section, Bibliographic Services Division (Figure 23, middle). Here the journals are given first to a highly trained clerk in the Index Section, who verifies the Journal Title Code and transliterates the title and names of authors for all journals printed in Cyrillic alphabets. This clerk also separates the journals into categories: those to be indexed in depth, the non-depth journals, those to be handled on a "rush" basis for processing, and those to be selectively indexed for medically related papers only. The journal issues then are

MEDLARS INPUT SUBSYSTEM FLOW CHART

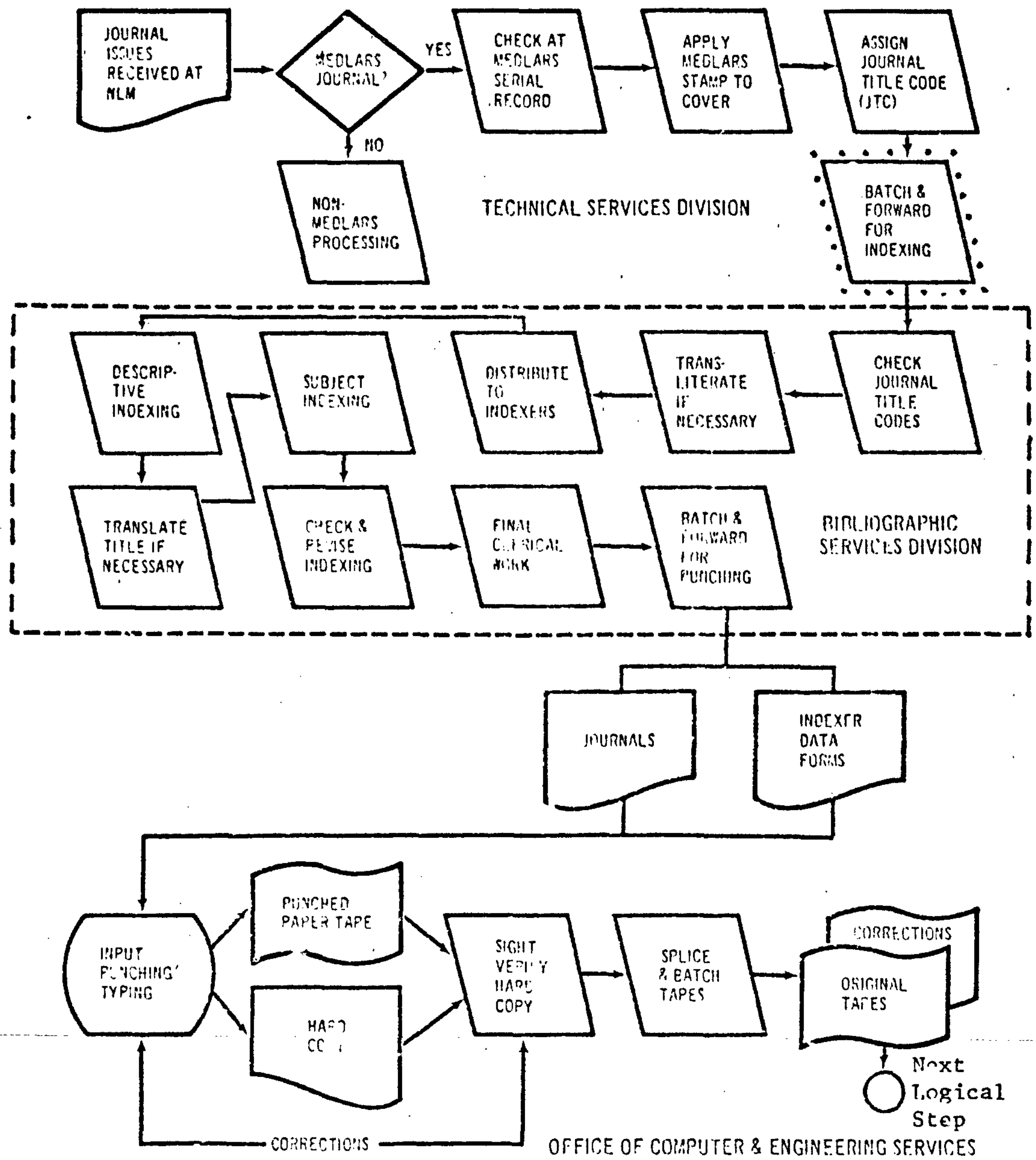


Figure 23

distributed to the professional indexers, taking into consideration the special subject or foreign-language skills of each individual.

The indexers prepare an Indexer Data Form for each article in the journal. The indexer first scans and evaluates the article to find out what it is about and what are the most important points to be covered. Subject headings and subheadings are assigned from the controlled vocabulary--MeSH. The Indexer Data Form includes several check tags which serve as reminders to the indexer of concepts which always are to be covered (e.g., age groups, clinical report, etc.). In handling a depth journal, the indexer may use as many subject headings as are needed to describe fully the content of the articles. When indexing a non-depth journal, the indexer is limited to subject headings that describe the primary concepts only. As of January 1968, depth journal articles were assigned an average of about 10 subject headings and non-depth journals were assigned an average of about 4. The indexer also assigns subheadings and must insure that he uses a valid main heading/subheading combination in each case that a subheading is used. In addition to assigning MeSH terms, the indexer decides whether each term is to be "print" or "non-print"; that is, to be printed in "Index Medicus" or to be used only in the retrieval process.

After indexing, the journals with data forms attached are sent to the revisers (senior professionals who check and revise the work of the indexers). After completion of work by the professional indexers and revisers, the journals go to a final clerical work station, where "sort authors" are established. Sort authors are required in cases where the computer is not able to follow its normal collating sequence in preparing alphabetic author list (e.g., St. Lawrence to sort as Saint Lawrence).

The original MEDLARS philosophy was to perform all indexing centrally with NLM staff. However, the massive volume of work to be done, coupled with rapidly increasing backlogs, caused Library management to reconsider this policy and begin to use outside contractors for some of the indexing work.[7] (See Figure 21, p. 131, Decentralized Indexing).

Decentralized indexing is now under way at such places as Keio University in Japan; the MEDLARS Stations at Harvard, the University of Alabama, and the University of Colorado; and in Israel, using PL 480 counterpart funds. Private contractors also have been used. This decentralized indexing has proven quite effective.[8] The "MEDLARS Indexing Manual" insures standardization of indexing and facilitates indexer training.

After completion of all Index Section tasks, batches of journals and data sheets are forwarded to the Office of Computer and Engineering Services for data punching and computer processing (Figure 23, bottom).

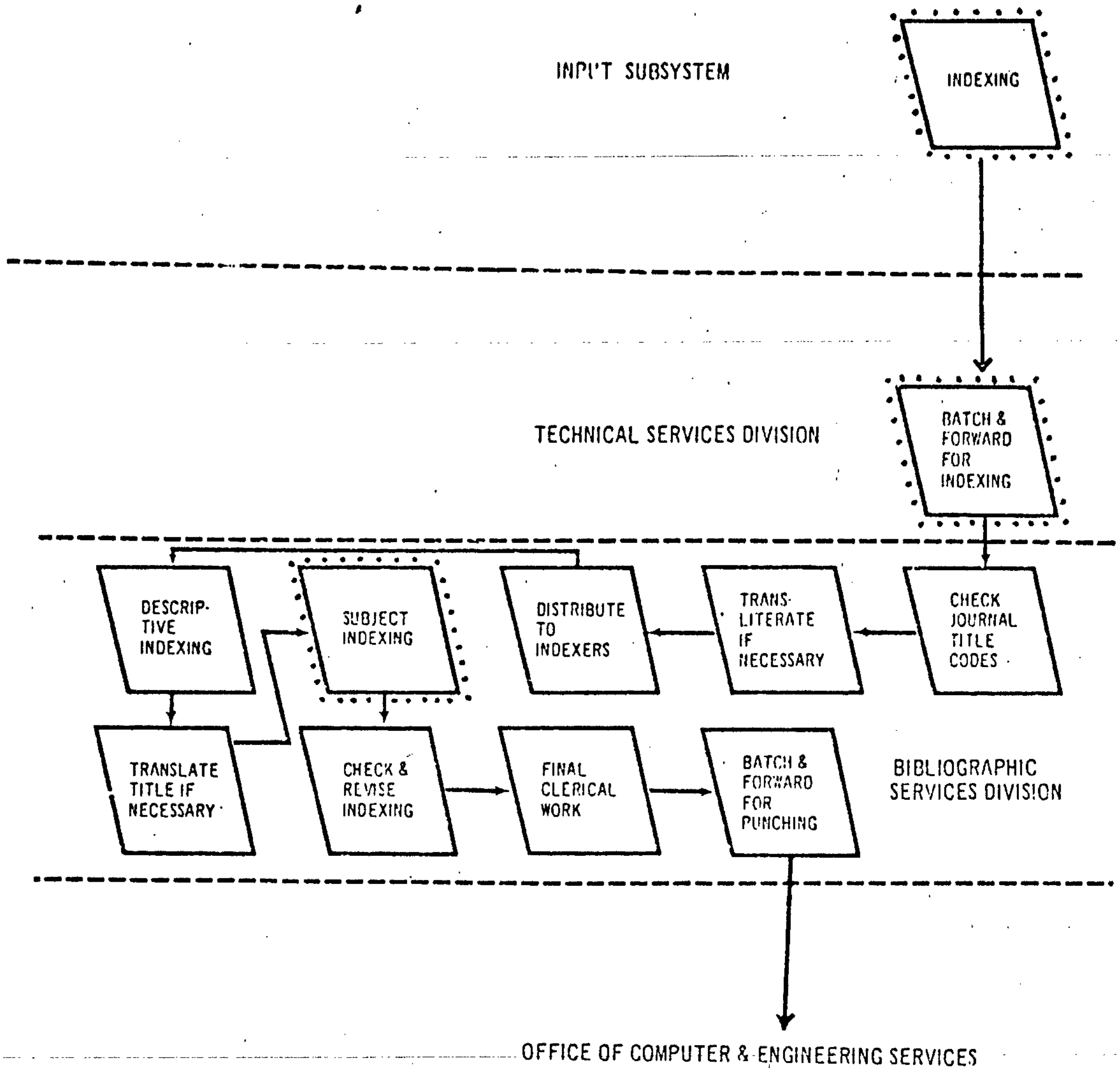
Figure 24 filters out those steps from Figures 22 and 23 which are beyond our scope and focuses on the subject indexing function of MEDLARS.

To recall, our purpose was to focus or gradually zero in on one of the network activities as a basic functional unit of the system. We started with the <u>Umbrella Net</u> of a hypothetical information system, and then switched to the <u>Umbrella Net</u> of our object system, MEDLARS. According to the design of our experiment, we then developed a PERT network of the subject indexing function of MEDLARS. From this function, we have selected "Indexing. Preparation of Indexer Data Forms," activity 8-9, Figure 21, for micromanipulation. Nevertheless, we can keep going

---

[7] <u>Ibid</u>., p. 20.

[8] <u>Ibid</u>.



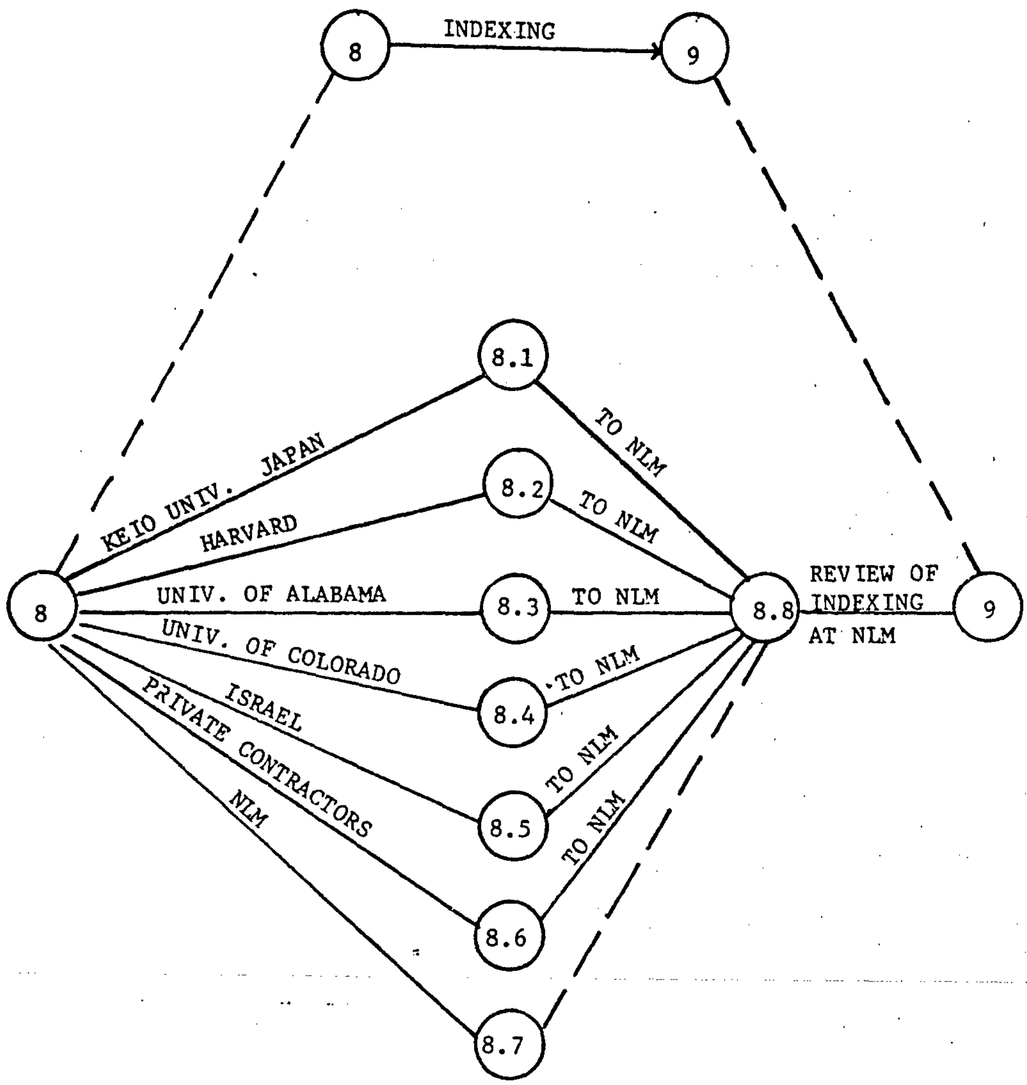
INDEXING FUNCTION OF MEDLARS

Figure 24

expanding the "family tree" as necessary to arrive at the basic functional units appropriate for the particular system being designed. For example, if we take activity 8-9 above, we will have to expand it like the following Figure 25. (For data see p.136).

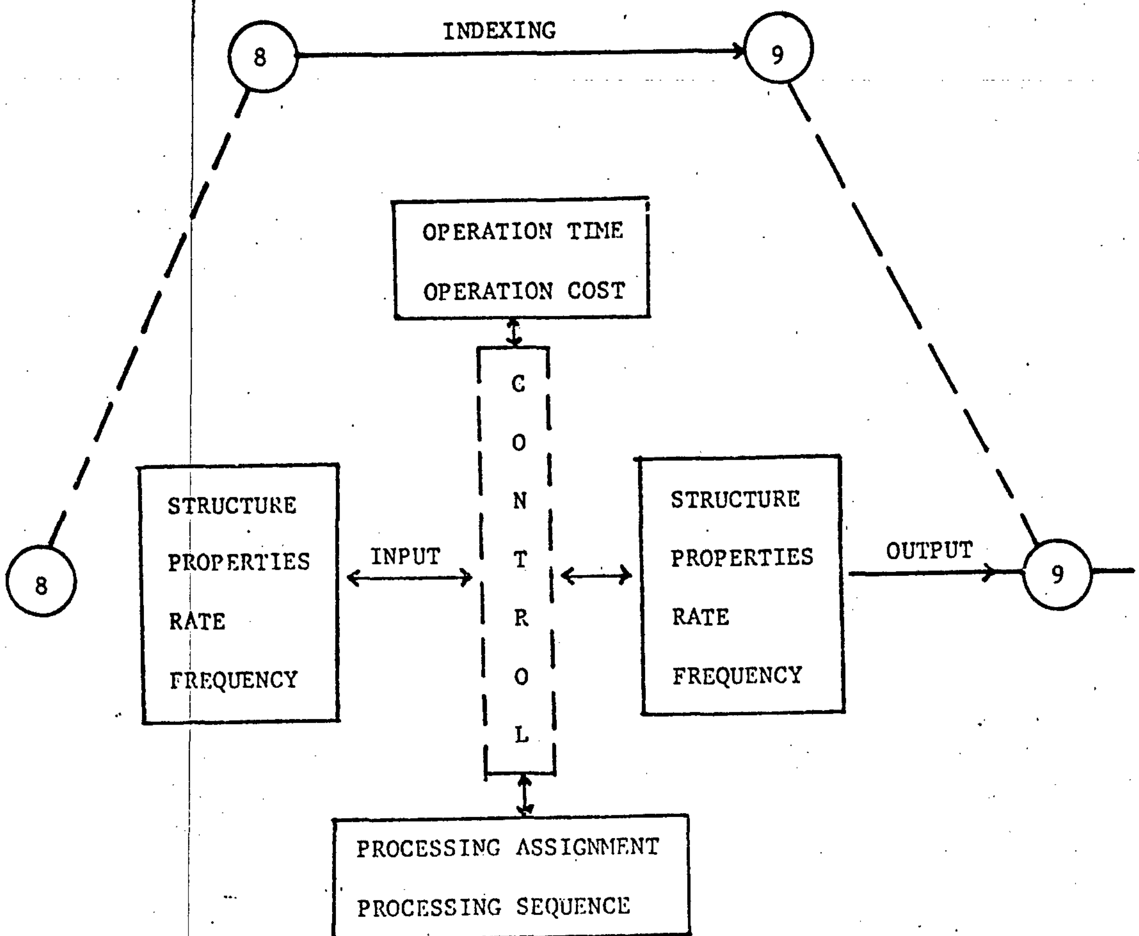
For the micromanipulation of the activities we have to refer to the discussion on the "Modified PERT Activity" (pp.118-20). Like any other in the network, activity 8-9 will involve everything as illustrated in the following Figure 26. As input, activity 8-9 will receive the journals at a certain rate and frequency; for example, 50 journals, twice a day. Structure and properties of this input relate to the journals as physical objects. The same is true with the output of this activity, and this is outside the scope of this study.

For computing indexing time and indexing cost (operation time and operation cost for any other activity), we have developed the PERT Computational Program, and adapted the CPM computational procedure. For indexing Assignment and indexing Sequence, we have adapted the Assignment Model and Sequencing Model respectively. The Program, Procedure, and the Models that follow this section are self-contained and self-explanatory units with proper examples and tutorials. A specific activity like the activity 8-9 can be routed through the general Program, Procedure, and Models to obtain the computed values.



EXPANSION OF ACTIVITY 8 - 9

Figure 25



MODIFIED PERT ACTIVITY 8-9

Figure 26

X. THE PERT COMPUTATIONAL PROGRAM

Term	Dasription	Referent
t_o	Optimistic time	Activity
t_m	Most likely time	Activity
t_p	Pessimistic time	Activity
t_E	Expected time	Activity
T_E	Earliest expected time	Event
T_L	Latest allowable time	Event
Slack	Project schedule time minus length of path	Path

The following PERT computational program has been written in PIL/L (Pitt Interpretive Language/50) for the IBM 360/50, to run on PTSS (Pitt Time Sharing System). It is a non-diagnostic, interactive mode computational program. It will accept input through the IBM 2741 terminal and provide output on the same.

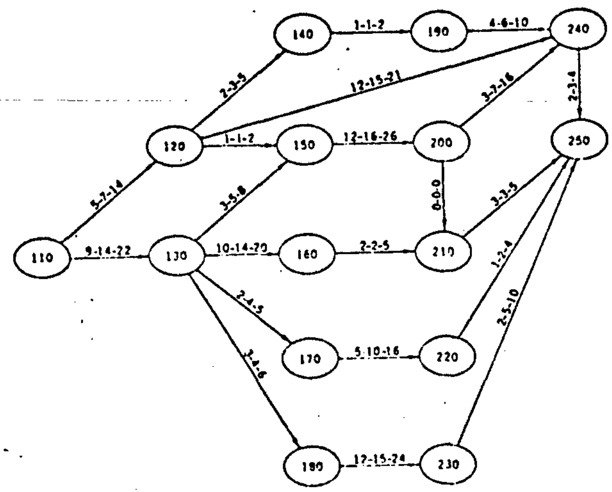
No special skill is necessary to run the program. Any secretary with an understanding of PERT terminology can work with the program to obtain PERT computational data and develop the necessary tables. This has been tested on some secretaries and found to be true.

The program is in six parts (part 3 not used). Part 1 gives some term and variable name explanations and states the equation for the calculation of the expected time for an activity. Part 1 automatically moves into Part 2, where the expected times of activities and events are calculated and printed in a tabular form. The program will not automatically move from here. The user has the option of either stopping here or moving on to the next part by typing "do part x," x being the part number. Parts 4, 5, and 6 calculate the "latest allowable time," "slack," and standard deviation, respectively.

Following is an illustrative example of the use of the program. The 21-activity network and data has been taken from Evarts.[1]

[1]Harry F. Evarts, Introduction to PERT (Boston: Allyn and Bacon, Inc., 1964), pp. 45-69, passim.

The user will start with the network with the three time estimates for each activity and event numbered sequentially, as shown below in Figure 27.



THE 21-ACTIVITY NETWORK

Figure 27

The user will also need a preliminary worksheet as shown below in Table 6.

Computation Worksheet								
Successor event	Predecessor event	t_o	t_m	t_p	t_E	T_E	T_L	Slack
250	240	2	3	4				
	230	2	5	10				
	220	1	2	4				
	210	3	3	5				
240	200	3	7	16				
	190	4	6	10				
	120	12	15	21				
230	180	12	15	24				
220	170	5	10	16				
210	200	0	0	0				
	160	2	2	5				
200	150	12	16	26				
190	140	1	1	2				
180	130	3	4	6				
170	130	2	4	5				
160	130	10	14	20				
150	130	3	5	8				
	120	1	1	2				
140	120	2	3	5				
130	110	9	14	22				
120	110	5	8	14				

TABLE 6

The worksheet has nine columns. The first five of these (successor event, predecessor event, and the three time-estimate columns) are filled in simply by recording the information from the network.

The first event recorded on the worksheet is the end event (250 in this case), which is placed at the top of the successor event column. Next, all the events immediately preceding event 250 are recorded in the predecessor event column, beginning with the highest numbered predecessor on the same line with event 250, and then on down the column until all immediate predecessors of event 250 are listed on separate lines. In this case, four predecessors--240, 230, 220, 210--are listed since these four events are directly connected to event 250 by activity arrows.

The next step is to return to the successor column to list the event numerically next lowest to event 250. Event 240, in this case, would be listed at this time as a successor event, and its three predecessors (found by tracking back the three arrows leading to event 240) would then be listed in the predecessor event column as described above. The three predecessors of event 240 are 200, 190, and 120.

The third number to appear in the successor event column is the next lowest number numerically of all those on the network. This is not necessarily the second successor event's highest numbered predecessor. In this case, the third event in the successor column is 230, which is not listed among 240's predecessors at all. In preparing worksheets, it is always important to refer back to the network for successor event numbers rather than to refer to the predecessor event column for this information.

This listing of events and their predecessors should proceed, with successor events in exact reverse numerical order, until the start event of the network is reached. Every event on the network, except the very first one, must appear in its proper order in the successor event column of the worksheet. Every event, except the very last one, must appear at least once in the predecessor event column, and many may appear more than once, although in no special numerical order.

After both the event columns are filled in and checked for order, the optimistic, most likely, and pessimistic times for each activity are taken from the network and put on the worksheet. On the worksheet the first line is for the times of activity 240-250, the second for activity 230-250, the third for activity 220-250, and so on. The tenth line is for activity 200-210, a dummy activity for which the times must be recorded even though they are simply 0-0-0. The last line is for activity 110-120, the first activity of the network. After all these events and activity times are recorded, the worksheet has been properly prepared for input to the PERT computational program.

As the program will demand the optimistic, most likely, and pessimistic times, the user will provide these values for each activity from the worksheet, but before that, the user will have to provide a value where the program will demand "n=>". This is to let the program know how many cycles it has to go through the computational loop before it can print the saved values in a tabular form. The user should provide a value which is equal to the number of necessary calculations. The user will keep providing the program the three time estimates for each activity as they are demanded by the program, by going down the worksheet until an expected time is calculated for each activity. The expected time (t_E) of an activity is calculated as follows:

$$t_E = \frac{t_o + 4t_m + t_p}{6}$$

The earliest expected time (T_E) to achieve an event is automatically computed by the program, upon receipt of the necessary values from the user. As soon as the user will provide the last set of values, the program will print in a tabular form the computed values of t_E and T_E. The following illustrations show the parts of the program (Parts 1 and 2) that will do this job, and output for values provided from the worksheet. The earliest expected time (T_E) for each event is calculated as follows:

$$T_E(\text{successor}) = T_E(\text{predecessor}) + t_E(\text{activity})$$

```
type part 1.

    1.001   type "You are now working with  BOSE/PERT  program."
    1.002   for i= 1 to 3: line.
    1.01    type "                 Te = Expected Time".
    1.02    type "                  o = Optimistic Time".
    1.03    type "                  p = Pessimistic Time".
    1.04    type "                  m = Most Likely Time".
    1.05    type "4= Weight of the Most Likely Time."
    1.06    for i=1 to 4: line.
    1.07    type "                        The Equation is".
    1.08    for i=1 to 2: line.
    1.09    type " Te = (o + 4*m + p)/ 6".
    1.1     for i=1 to 4: line.
    1.11    do part 2.

>
type part 2.

    2.02    for counter= 0: set i= 0.
    2.022   demand n.
    2.024   set a= "Te= ___ .__".
    2.025   set d= "     _: __   .__             ___  : ____.___"
    2.026   set e= "_____: ___  .__".
    2.028   demand o, m, p.
    2.042   for i= i+1: for Te=(o+4*m+p)/6: set ActT(i)= Te.
    2.044   type in form a, Te.
    2.046   set counter= counter+1.
    2.048   type counter.
    2.062   if counter= n, to step 2.064; to step 2.028.
    2.064   line.
    2.066   type "Activity Time".
    2.068   line.
    2.082   for i= 1 to i: type in form e, i, ActT(i).
    2.084   type "Done".
    2.086   type "Calculation of Event Time".
    2.088   for counter= 0: set k= n+1.
    2.089   type "When no more te to add, type 0 (zero) when te is demanded."
    2.092   type "te= expected time of an activity".
    2.093   set SumTe= 0.
    2.094   demand te.
    2.096   set SumTe= SumTe+te.
    2.098   if te= 0, to step 2.221; to step 2.094.
    2.221   for k= k-1: set Evnt(k)= SumTe.
    2.222   set counter= counter+1.
    2.223   type SumTe.
    2.224   type counter.
    2.226   if counter= n, to step 2.228; to step 2.093.
    2.228   type "     Activity Time                  Event Time"
    2.242   for i= 1 to n: type in form d, i, ActT(i), i, Evnt(i).
    2.244   type "Done".
    2.246   Done.

>
```

```
>do part 1.

You are now working with  BOSE/PERT  program.


               Te = Expected Time
                o = Optimistic Time
                p = Pessimistic Time
                m = Most Likely Time
4= Weight of the Most Likely Time.


                  The Equation is

 Te = (o + 4*m + p)/ 6
```

```
  n =>21

 o =>2
 m =>3
 p =>4
 Te=    3.00
 counter =  1.0
 o =>2
 m =>5
 p =>10
 Te=    5.33
 counter =  2.0
 o =>1
 m =>2
 p =>4
 Te=    2.16
 counter =  3.0
 o =>3
 m =>3
 p =>5
 Te=    3.33
 counter =  4.0
 o =>3
 m =>7
 p =>16
 Te=    7.83
 counter =  5.0
 o =>4
 m =>6
 p =>10
 Te=    6.33
 counter =  6.0
 o =>12
 m =>15
 p =>21
 Te=   15.50
 counter =  7.0
 o =>12
 m =>15
 p =>24
 Te=   16.00
 counter =  8.0
 o =>5
 m =>10
 p =>16
 Te=   10.16
 counter =  9.0
 o =>0
 m =>0
 p =>0
 Te=    0.00
 counter =  10.0
```

```
o =>2
m =>2
p =>5
Te=     2.50
counter =  11.0
o =>12
m =>16
p =>26
Te=   17.00
counter =  12.0
o =>1
m =>1
p =>2
Te=     1.16
counter =  13.0
o =>3
m =>4
p =>6
Te=     4.16
counter =  14.0
o =>2
m =>4
p =>5
Te=     3.83
counter =  15.0
o =>10
m =>14
p =>20
Te=   14.33
counter =  16.0
o =>3
m =>5
p =>8
Te=     5.16
counter =  17.0
o =>1
m =>1
p =>2
Te=     1.16
counter =  18.0
o =>2
m =>3
p =>5
Te=     3.16
counter =  19.0
o =>9
m =>14
p =>22
Te=   14.50
counter =  20.0
o =>5
m =>8
p =>14
Te=     8.50
counter =  21.0
```

Activity Time

1:	3.00
2:	5.33
3:	2.16
4:	3.33
5:	7.83
6:	6.33
7:	15.50
8:	16.00
9:	10.16
10:	0.00
11:	2.50
12:	17.00
13:	1.16
14:	4.16
15:	3.83
16:	14.33
17:	5.16
18:	1.16
19:	3.16
20:	14.50
21:	8.50

Done

```
Calculation of Event Time
When no more te to add, type 0 (zero) when te is demanded.
te= expected time of an activity
te =>8.5
te =>0
SumTe =  8.5
counter =  1.0
te =>14.5
te =>0
SumTe =  14.5
counter =  2.0
te =>8.5
te =>3.2
te =>0
SumTe =  11.7
counter =  3.0
te =>8.5
te =>1.2
te =>0
SumTe =  9.7
counter =  4.0
te =>14.5
te =>5.2
te =>0
SumTe =  19.7
counter =  5.0
te =>14.5
te =>14.3
te =>0
SumTe =  28.8
counter =  6.0
te =>14.5
te =>3.8
te =>0
SumTe =  18.3
counter =  7.0
te =>14.5
te =>4.2
te =>0
SumTe =  18.7
counter =  8.0
te =>11.7
te =>1.2
te =>0
SumTe =  12.9
counter =  9.0
te =>19.7
te =>17.0
te =>0
SumTe =  36.7
counter =  10.0
```

```
te =>28.8
te =>2.5
te =>0
SumTe =  31.3
counter =  11.0
te =>36.7
te =>0
SumTe =  36.7
counter =  12.0
te =>18.3
te =>10.2
te =>0
SumTe =  28.5
counter =  13.0
te =>18.7
te =>16.0
te =>0
SumTe =  34.7
counter =  14.0
te =>8.5
te =>15.5
te =>0
SumTe =  24.0
counter =  15.0
te =>12.9
te =>6.3
te =>0
SumTe =  19.2
counter =  16.0
te =>36.7
te =>7.8
te =>0
SumTe =  44.5
counter =  17.0
te =>36.7
te =>3.3
te =>0
SumTe =  40.0
counter =  18.0
te =>28.5
te =>2.2
te =>0
SumTe =  30.7
counter =  19.0
te =>34.7
te =>5.3
te =>0
SumTe =  40.0
counter =  20.0
te =>44.5
te =>3.0
te =>0

SumTe =  47.5
counter =  21.0
```

```
   Activity Time                     Event Time
  1:      3.00                       1:    47.50
  2:      5.33                       2:    40.00
  3:      2.16                       3:    30.70
  4:      3.33                       4:    40.00
  5:      7.83                       5:    44.50
  6:      6.33                       6:    19.20
  7:     15.50                       7:    24.00
  8:     16.00                       8:    34.70
  9:     10.16                       9:    28.50
 10:      0.00                      10:    36.70
 11:      2.50                      11:    31.30
 12:     17.00                      12:    36.70
 13:      1.16                      13:    12.90
 14:      4.16                      14:    18.70
 15:      3.83                      15:    18.30
 16:     14.33                      16:    28.80
 17:      5.16                      17:    19.70
 18:      1.16                      18:     9.70
 19:      3.16                      19:    11.70
 20:     14.50                      20:    14.50
 21:      8.50                      21:     8.50
Done
>
```

In the preceding example, T_E for event 120 is 8.5 (number in the table is 21), the expected time for completion of the activity 110-120. Similarly, T_E for event 160 is 14.5 plus 14.3 (t_E for the activity 130-160), a total of 28.8. When a successor event has more than one activity arrow leading to it, the user will calculate more than one T_E. The greatest should be circled and used in calculating T_E for succeeding activities. For example, T_E for event 150 is 19.7 rather than 9.7. Therefore, T_E for event 200 is 19.7 + 17.0, or 36.7. The purpose of using the greatest number for subsequent calculations is to assure that enough time is allowed for the path consuming the greatest amount of time.

Part 4 of the program will compute the latest allowable time (T_L) which refers to the time by which an event must be completed if the project is to be completed on schedule.

T_L for any event is calculated by subtracting from the scheduled length of the project the length of the longest path backward from the end of the network to the event in question. In those instances in which a project does not have a scheduled completion time, the T_E of the end event is also used as T_L for that event. For example, if the scheduled time for event 250 is 45.0, then the latest allowable time for event 250, designated as T_L in the worksheet, is 45.0. The latest allowable time for the predecessor of event 250 is calculated as follows:

$$T_L(\text{predecessor}) = T_L(\text{successor}) - t_E(\text{activity})$$

Thus, for event 240, T_L equals 45.0 (T_L for event 250) minus 3.0 (t_E for the activity 240-250), or 42.0. When an event has two or more succeeding activities, more than one T_L figure will be calculated. The lowest of these figures should be used. For example, event 200 appears twice in the predecessor event column. For successor event 240, T_L for event 200 is 42.0 - 7.8 which equals 34.2. For successor event

210, T_L for event 200 is 41.7 - 0.0, which equals 41.7. The lower figure, 34.2, should be used since this will assure that enough time is allowed for the path consuming the greatest amount of time.

The following illustrations show the parts of the program (Part 4) that will compute the latest allowable time by which an event must be completed to meet the schedule, for values provided from the worksheet.

type part 4.

```
4.001    type "Calculation of Latest Allowable Time by which"
4.002    type "an event must be completed.".
4.003    for Counter=0: set j=0.
4.004    demand n.
4.006    set e="  ___: ___ .__".
4.01     type "Ltest=Latest Allowable Time.".
4.02     type "SumTe=STe and Te=ExpT.".
4.021    demand STe, ExpT.
4.03     set Ltest=STe-ExpT.
4.04     set c="Ltest=__  .__".
4.05     line.
4.06     type in form c, Ltest.
4.061    for j=j+1: set Latest(j)=Ltest.
4.07     set Counter=Counter+1.
4.08     type Counter.
4.09     if Counter=n, to step 4.101.
4.091    for i=1 to 2: line.
4.1      to step 4.021.
4.101    line.
4.11     type "  Latest Allowable Time".
4.111    line.
4.113    for i=1 to j: type in form e, i, Latest(i).
4.12     type "Done.".
4.13     Done.
```

```
do part 4.

Calculation of Latest Allowable Time by which
an event must be completed.
n =>20
Ltest=Latest Allowable Time.
SumTe=STe and Te=ExpT.
STe =>45

ExpT =>0

Ltest=  45.00
Counter =   1.0


STe =>45
ExpT =>3

Ltest=  42.00
Counter =   2.0


STe =>45
ExpT =>5.3

Ltest=  39.70
Counter =   3.0


STe =>45
ExpT =>2.2

Ltest=  42.80
Counter =   4.0


STe =>45
ExpT =>3.3

Ltest=  41.70
Counter =   5.0


STe =>42.0
ExpT =>7.8

Ltest=  34.20
Counter =   6.0


STe =>41.7
ExpT =>0

Ltest=  41.70
Counter =   7.0
```

```
STe =>42
FxpT =>6.3

Ltest=  35.70
Counter =  8.0


STe =>39.7
ExpT =>16

Ltest=  23.70
Counter =  9.0


STe =>42.8
ExpT =>10.2

Ltest=  32.60
Counter =  10.0


STe =>41.7
ExpT =>2.5

Ltest=  39.20
Counter =  11.0


STe =>34.2
ExpT =>17.0

Ltest=  17.20
Counter =  12.0


STe =>35.7
ExpT =>1.2

Ltest=  34.50
Counter =  13.0


STe =>23.7
ExpT =>4.2

Ltest=  19.50
Counter =  14.0
```

```
STe =>32.6
ExpT =>3.8

Ltest=  28.80
Counter =  15.0


STe =>39.2
ExpT =>14.3

Ltest=  24.90
Counter =  16.0


STe =>17.2
ExpT =>5.2

Ltest=  12.00
Counter =  17.0


STe =>17.2
ExpT =>1.2

Ltest=  16.00
Counter =  18.0


STe =>42
ExpT =>15.5

Ltest=  26.50
Counter =  19.0


STe =>34.5
ExpT =>3.2

Ltest=  31.30
Counter =  20.0
```

```
        Latest Allowable Time

           1:    45.00
           2:    42.00
           3:    39.70
           4:    42.80
           5:    41.70
           6:    34.20
           7:    41.70
           8:    35.70
           9:    23.70
          10:    32.60
          11:    39.20
          12:    17.20
          13:    34.50
          14:    19.50
          15:    28.80
          16:    24.90
          17:    12.00
          18:    16.00
          19:    26.50
          20:    31.30
Done.
>
```

The slack of a particular path is the difference between the time scheduled for the entire project and that needed for the path. The slack of a path is positive if the time at which the final event of the path is expected to be achieved occurs earlier than the project completion date. If the time is later than the completion date, the slack is negative. Paths with negative slack become critical paths, the one with the greatest negative figure being the critical path (that is, longest and needing most attention). The calculation is as follows:

$$\text{Slack} = T_L - T_E$$

```
type part 5.

    5.002   set j=0.
    5.004   demand n.
    5.006   set e="  __ __: ___. .__".
    5.01    set Counter=0.
    5.02    type "Slack=Ltest-SumTe".
    5.021   line.
    5.03    type "Slack is the difference between"
    5.031   type "the Expected Events Completion Time"
    5.04    type "and the Latest Allowable Time for that Event."
    5.041   type "Ltest=Ltst, and STe=SmTe.".
    5.05    for i=1 to 3: line.
    5.06    demand Ltst, SmTe.
    5.07    set Slack=Ltst-SmTe.
    5.072   for j=j+1: set Slck(j)=Slack.
    5.08    set d="Slack=__. _.__".
    5.09    for i=1 to 3: line.
    5.1     type in form d, Slack.
    5.11    set Counter=Counter+1.
    5.12    type Counter.
    5.13    if Counter = n, to step 5.132; to step 5.06.
    5.132   type "      SLACK".
    5.14    for i=1 to j: type in form e, i, Slck(i).
    5.15    for i=1 to 3: line.
    5.16    type "Done."
    5.17    Done.

>
```

```
do part 5.

n =>21
Slack=Ltest-SumTe

Slack is the difference between
the Expected Events Completion Time
and the Latest Allowable Time for that Event.
Ltest=Ltst, and STe=SmTe.


Ltst =>45
SmTe =>47.5


Slack=  -2.50
Counter =  1.0
Ltst =>45
SmTe =>40


Slack=   5.00
Counter =  2.0
Ltst =>45
SmTe =>30.7


Slack=  14.30
Counter =  3.0
Ltst =>45
SmTe =>40


Slack=   5.00
Counter =  4.0
Ltst =>42
SmTe =>44.5


Slack=  -2.50
Counter =  5.0
Ltst =>42
SmTe =>19.2
```

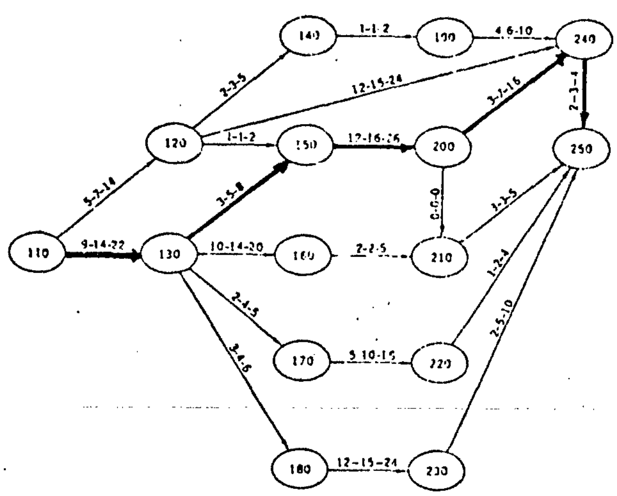
```
Slack=  22.80
Counter =   6.0
Ltst =>42
SmTe =>24

Slack=  18.00
Counter =   7.0
Ltst =>39.7
SmTe =>34.7

Slack=   5.00
Counter =   8.0
Ltst =>42.8
SmTe =>28.5

Slack=  14.30
Counter =   9.0
Ltst =>41.7
SmTe =>36.7

Slack=   5.00
Counter =  10.0
Ltst =>41.7
SmTe =>31.3

Slack=  10.40
Counter =  11.0
Ltst =>54.2
SmTe =>36.7

Slack=  -2.50
Counter =  12.0
Ltst =>35.7
SmTe =>12.9
```

```
Slack=   22.80
Counter =   13.0
Ltst =>23.7
SmTe =>18.7

Slack=    5.00
Counter =   14.0
Ltst =>32.6
SmTe =>18.3

Slack=   14.30
Counter =   15.0
Ltst => 39.2
SmTe =>28.8

Slack=   10.40
Counter =   16.0
Ltst =>17.2
SmTe =>19.7

Slack=   -2.50
Counter =   17.0
Ltst =>17.2
SmTe =>9.7

Slack=    7.50
Counter =   18.0
Ltst =>34.5
SmTe =>11.7

Slack=   22.80
Counter =   19.0
Ltst =>12
SmTe =>14.5

Slack=   -2.50
Counter =   20.0
Ltst =>16
SmTe =>8.5
```

```
Slack=    7.50
Counter =   21.0
     SLACK
    1:    -2.50
    2:     5.00
    3:    14.30
    4:     5.00
    5:    -2.50
    6:    22.80
    7:    18.00
    8:     5.00
    9:    14.30
   10:     5.00
   11:    10.40
   12:    -2.50
   13:    22.80
   14:     5.00
   15:    14.30
   16:    10.40
   17:    -2.50
   18:     7.50
   19:    22.80
   20:    -2.50
   21:     7.50


Done.
>
```

The objective of all these calculations is to identify the critical path, the semicritical paths, and the slack paths. The critical path begins with start event, terminates with end event, and lies along those activities which show the identical slack figure which is the lowest positive figure or the greatest negative figure.

In our example, a negative figure (-2.5) appears in the worksheet. By beginning at the bottom of the slack column of the worksheet and working up to find the first -2.5 slack, the analyst can identify the critical path by jotting down both the predecessor event and the successor event on the same line with the first -2.5, then the successor events of each -2.5 slack line on the worksheet. The critical path in this case would be 110-130-150-200-240-250. The heavy lines on the network show the critical path in Figure 28.



NETWORK SHOWING THE CRITICAL PATH

Figure 28

Slack affects equally an entire path, not just one activity. For instance, the critical path slack of -2.5 refers to the entire path. If the time for activity 150-200 (17 weeks) could be reduced to 14.5 weeks, the -2.5 slack for the entire path would be canceled and the slack would become zero.

Selection of semicritical and slack paths in a network, after the critical path is identified, is a matter of judgment. Selection of these paths must depend on arbitrary decisions about time, since nothing else of the project is known.

Semicritical paths in this case are:

Path	Slack
130-180-230-250	5.0
200-210-250	5.0
110-120-150	7.5

Slack paths are:

Path	Slack
130-160-210	10.4
130-170-220-250	14.3
120-240	18.0
120-140-190-240	22.8

Part 6 of the program will determine the probability of completion of the events and the project on schedule. Usually a project has a scheduled completion date, and it is unlikely that such a date would coincide with the earliest expected time of the end event.

The project in the example is scheduled to be completed in 45 weeks. T_E for the last event in the project is 47.5 weeks. The probability that the project will be completed on time (i.e., 45 weeks) is less than .5 and is calculated statistically. The program will measure the σ from the mean in our example. The σ figure can be referred to the Table 8 which converts the deviation to a measure of the area under the normal curve beyond the scheduled date. The equation is as follows:

$$\sigma = \frac{T_S - T_E}{\sigma \Sigma \sigma^2}$$

in which:

T_S = scheduled completion time of the project

$\Sigma \sigma^2$ = the sum of the variances of the activities on the path being considered

$\sigma \Sigma \sigma^2$ = the standard deviation of the sum of the variances

We have to know the sum of the variances of those activities on the critical path. In order to determine the variance (σ^2), the following formula is used:

$$\sigma^2 = \left(\frac{t_p - t_o}{6}\right)^2$$

to find the variance of each activity. These variances are then totaled to give $\Sigma\sigma^2$. For ease in calculation, a table is constructed for the critical path as shown in Table 7.

Critical Path Activities					
Successor event	Predecessor event	t_o	t_p	$t_p - t_o$	$(t_p - t_o)^2$
250	240	2	4	2	4
240	200	3	16	13	169
200	150	12	26	14	196
150	130	3	8	5	25
130	110	9	22	13	169
				Total	563

TABLE 7

Following is the Program (Part 5) and the computational results:

```
type part 6.


    6.002   demand n.
    6.004   for i=0: for k=0: set Counter=0.
    6.006   set d="  _ : __  .__              ____: ____.__".
    6.01    type "Calculation of the Standard Deviation.".
    6.02    set SumDSQ=0.
    6.03    demand p, o.
    6.05    set DIF=p-o.
    6.051   type DIF.
    6.053   for j=j+1: set DIFF(j)=DIF.
    6.06    set DifSQR=(p-o)**2.
    6.061   type DifSQR.
    6.063   for k=k+1: set DIFFSQR(k)=DifSQR.
    6.07    set SumDSQ=SumDSQ+DifSQR.
    6.071   type SumDSQ.
    6.0711  for i=1 to 2: line.
    6.0713  set Counter=Counter+1.
    6.0715  type Counter.
    6.0717  if Counter= n, to step 6.072; to step 6.03.
    6.072   type "     DIFF                           DIFFSQR".
    6.074   line.
    6.076   for i=1 to j: type in form d, i, DIFF(i), i, DIFFSQR(i).
    6.078   line.
    6.08    type "SVRnc means Sum of the Variances.".
    6.09    set SVrnc=SumDSQ/(6**2).
    6.1     set DVrnc=SQRT of SVrnc.
    6.11    type "SKEDUL means Scheduled Completion Time of the Project.".
    6.12    type "StDev means Standard Deviation.".
    6.13    demand SKEDUL.
    6.132   demand SumTe.
    6.14    set StDev = (SKEDUL - SumTe)/ DVrnc.
    6.15    set e="StDev= ____.__".
    6.16    type in form e, StDev.
    6.17    type "Done.".
    6.18    Done.

>
```

```
do part 6.

n =>5
Calculation of the Standard Deviation.
p =>4
o =>2
DIF =  2.0
DifSQR =  4.0
SumDSQ =  4.0

Counter =  1.0
p =>16
o =>3
DIF =  13.0
DifSQR =  169.0
SumDSQ =  173.0

Counter =  2.0
p =>26
o =>12
DIF =  14.0
DifSQR =  196.0
SumDSQ =  369.0

Counter =  3.0
p =>8
o =>3
DIF =  5.0
DifSQR =  25.0
SumDSQ =  394.0

Counter =  4.0
p =>22
o =>9
DIF =  13.0
DifSQR =  169.0
SumDSQ =  563.0

Counter =  5.0
```

DIFF	
1:	2.00
2:	13.00
3:	14.00
4:	5.00
5:	13.00

DIFFSQR	
1:	4.00
2:	169.00
3:	196.00
4:	25.00
5:	169.00

```
SVRnc means Sum of the Variances.
SKEDUL means Scheduled Completion Time of the Project.
StDev means Standard Deviation.
SKEDUL =>45
SumTe =>47.5
StDev=    -0.63
Done.
>
```

The figure StDev = -0.63 refers to the number of deviations from the mean (T_E) to the scheduled date (T_S). By referring to the Table 8 the σ is converted to the percentage of area under the curve beyond T_S. The figure -0.63 is between -0.6 and -0.7, and by approximation we can determine that -0.63 is .26. That is, 26 percent of the area under the curve is to the left of T_S; so, there is a 26 percent probability that the project will be completed by the scheduled date.

Table of Normal Distribution

Normal deviate	Area	Normal deviate	Area
−0.0	.50	0.0	.50
−0.1	.46	0.1	.54
−0.2	.42	0.2	.58
−0.3	.38	0.3	.62
−0.4	.34	0.4	.66
−0.5	.31	0.5	.69
−0.6	.27	0.6	.73
−0.7	.24	0.7	.76
−0.8	.21	0.8	.79
−0.9	.18	0.9	.82
−1.0	.16	1.0	.84
−1.1	.14	1.1	.86
−1.2	.12	1.2	.88
−1.3	.10	1.3	.90
−1.4	.08	1.4	.92
−1.5	.07	1.5	.93
−1.6	.05	1.6	.95
−1.7	.04	1.7	.96
−1.8	.04	1.8	.96
−1.9	.03	1.9	.97
−2.0	.02	2.0	.98
−2.1	.02	2.1	.98
−2.2	.01	2.2	.99
−2.3	.01	2.3	.99
−2.4	.01	2.4	.99
−2.5	.01	2.5	.99

TABLE 8

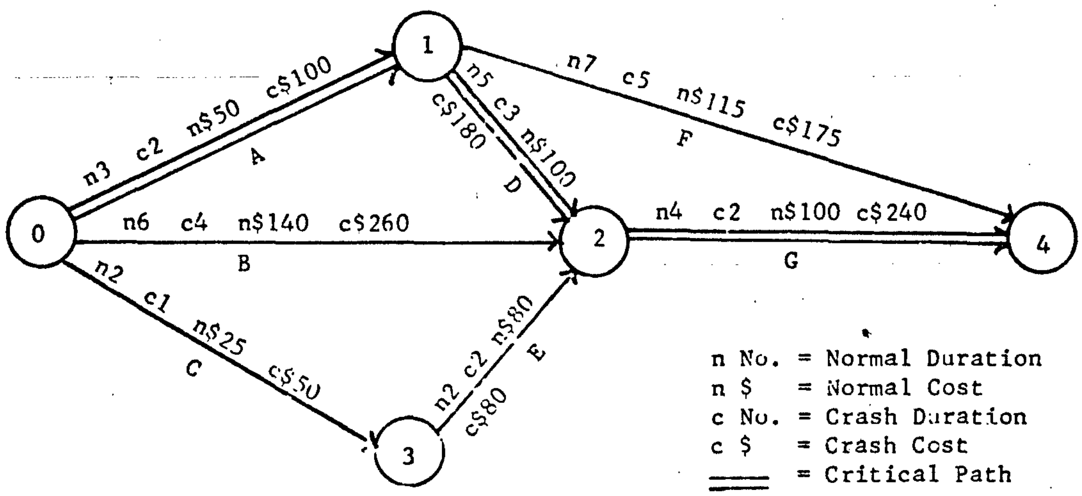
Probability values of .25 to .30 at the low end of the scale and .60 to .65 at the high end generally indicate the acceptable range of probability. When the calculated probability is below .25 or .30, the likelihood of meeting the project's scheduled completion date is so low that critical path time must be shortened. When probabilities are above .60 or .65, there is a strong likelihood that the project completion date will be met. In case of very high probability, management should consider using some of the resources committed to the project elsewhere in the system.

The critical path is the chain of activities through the project network with the longest duration between the beginning and the end of the project. This path of activities through the network determines the minimal (critical) time to complete the complex dependent set of activities. A change of time to complete any of the activities in the critical path will likewise change the total project duration.

Each activity is assigned a duration range and related cost. Each one of these various project durations produces different project costs. In the scheduling phase, the mathematics of CPM is used to compute these various project durations, and the lowest possible cost for each different project duration, thus producing the optimum schedule.

XI. CPM COMPUTATIONAL PROCEDURE[1]

The following network, Figure 29, gives normal and crash time/cost estimates for each activity. This information has been tabulated in Table 9 below.



NORMAL AND CRASH TIME/COST ESTIMATES

Figure 29

Activities	Normal Days	Normal Dollars	Crash Days	Crash Dollars
A	3	$ 50	2	$ 100
B	6	140	4	260
C	2	25	1	50
D	5	100	3	180
E	2	80	2	80
F	7	115	5	175
G	4	100	2	240
	Total	$ 610	Total	$ 1085

TABLE 9

COST ESTIMATE TABLE

[1] The Network and data have been adapted from Zalokar, op. cit., pp. 7-10, passim.

By referring to the above network, Figure 29, we see that the longest project duration using normal time estimates would be 12 days (or any other time unit), by following the critical path A, D, and G (double lines). The only way the project's duration can be reduced is to reduce the time of any of the activities falling on the critical path. Since in PERT/CPM it is assumed that cost is directly proportional to the time required for an activity, we have to make sure that the time reduction is made at the lowest possible cost. For this we need another piece of information for each activity--the activity cost slope or cost/time unit reduction. Using activity B as an example and assuming a linear relationship, the normal and crash estimates are presented graphically to illustrate the cost slope, in Figure 30, below.[2]

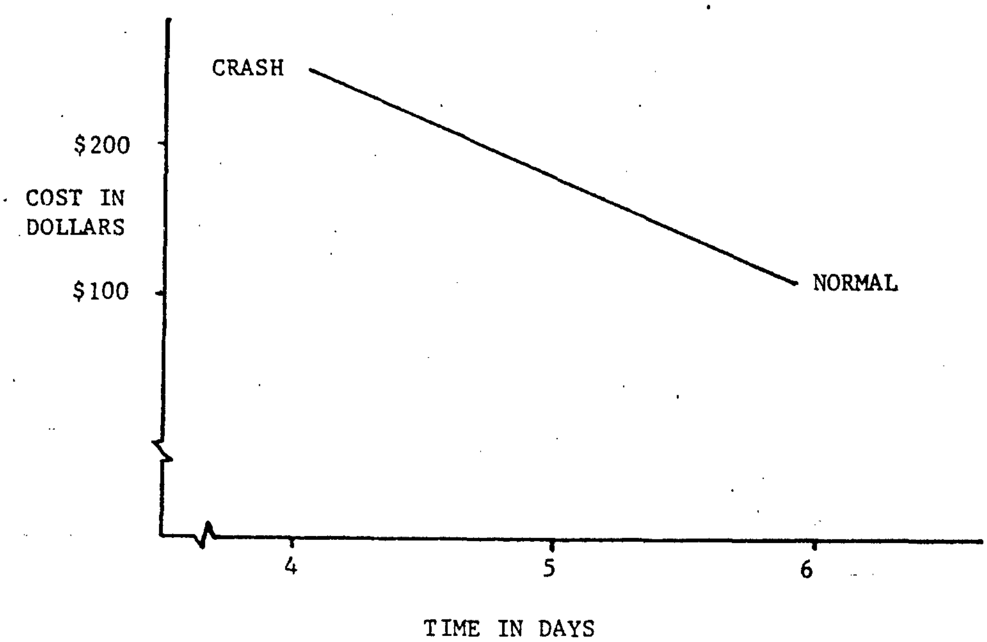


Figure 30

[2]Ibid., p. 9.

The cost slope of this curve is computed by the formula:

$$\frac{\text{Crash Cost - Normal Cost}}{\text{Normal Time - Crash Time}}$$

Substituting the respective values for the activity B from Table 9 we get

$$\frac{\$260 - \$140}{6 \text{ days} - 4 \text{ days}} = \frac{\$120}{2 \text{ days}} = \underline{\$60/\text{day}}$$

Computing this way for each activity, an additional column is added to Table 9 to produce the following Table 10.

Activities	Normal		Crash		Cost
	Days	Dollars	Days	Dollars	Slope
A	3	$ 50	2	$ 100	$ 50
B	6	140	4	260	60
C	2	25	1	50	25
D	5	100	3	180	40
E	2	80	2	80	--
F	7	115	5	175	30
G	4	100	2	240	70
	Total	$610	Total	$1085	

COST TABLE

TABLE 10

We see from Table 10 above that the 12-day normal duration of the project costs $610. The least expensive way to reduce the project duration by one day would be to reduce the time for activity D by one day, for an additional cost of $40, raising the project cost to $650. It can be easily seen by referring to the above Table 10 that reducing the time of the other activities on the critical path, activities A or G, would be more costly.

We may proceed this way until other paths become critical or reducing time of other activities become less expensive. It is important to take into consideration the marginal costs underlying the direct and indirect cost in the development of real-life schedules.

XII. THE ASSIGNMENT MODEL[1]

The problem of Assignment and Sequencing has relevance to both design and operation of information systems. These two operations research techniques have been applied in the following two models to indicate their applicability and facilitate their use in information system design and operation.

The problem of assignment is essentially a problem of optimum allocation of resources. In any situation where some product and/or services are being made available, there would be no problem of allocation had there been enough of all the necessary factors of production--land, labor, capital, and organization. But limitations of either the amount of the factors of production or the way they can be employed prevent us from having an ideal employment of the factors of production. In such a situation, we wish to allocate our available resources to the activities that will optimize the total return and effectiveness.

In assignment problems with a finite number of choices, we could, in theory, enumerate all possible choices, but in most cases enumeration will be too unwieldy; for example, there will be 10! ways of assigning, for subject analysis, 10 documents one apiece to 10 subject analysts. The technique of linear programming is used to analyze these situations. For the solution of an assignment problem of the nature we are talking about, n items are distributed among n boxes, one item to a box, in such a way that the return obtained from the distribution is optimized. Formally stated, the problem is: given an n-by-n array of real numbers (C_{ij}),

[1]Maurice Sasiene, Arthur Yaspan, and Lawrence Friedman, Operations Research--Methods and Problems, Wiley International Edition (New York: John Wiley & Sons, Inc., 1959), pp. 183-192, passim.

where C_{ij} is the individual return associated with assigning the i th item to the j th box, to find among all permutations $(i_1, i_2, \ldots, i_n)$, of the set of integers (1, 2, ..., n), that permutation for which

$$C_{1i_1} + C_{2i_2} + \ldots + C_{ni_n}$$

takes its maximum (minimum) value.

There are n! ways of assigning n items to n boxes. The following example illustrates the method of choosing the optimal permutation or assignment.

A subject-analysis department head in an information center has four subject analysts, and four documents to be analyzed. The analysts differ in efficiency and depth of subject knowledge, and the documents differ in sophistication of treatment and depth. His estimates of the times each analyst would take to perform each document-analysis is given in the effectiveness matrix below. The problem is: how should the job be assigned, one to an analyst, so as to minimize the total man-hours.

ANALYSTS

DOCUMENTS	I	II	III	IV
A	8	25	17	11
B	13	28	4	26
C	38	19	18	15
D	19	26	24	10

THE EFFECTIVENESS MATRIX

TABLE 11

There are 4! possible sets of associations that satisfy these conditions. All the possible sets can be written down, together with the corresponding total man-hours, but the more systematic approach is to take the smallest number in row A and subtract it from each element in the row. The result in our example is:

	I	II	III	IV
A	0	18	9	3
B	13	28	4	25
C	38	19	18	15
D	19	26	24	10

THE EFFECTIVENESS MATRIX

TABLE 12

Assuming we have assigned one analysis job to each analyst, no matter whatever assignment we have made, the total man-hours for the new matrix will be 8 less than for the old matrix. Hence an assignment that minimizes the total for one matrix also minimizes the total for the other. The basis for the solution is the theorem: "If in an assignment problem we add a constant to every element of a row (or column) in the effectiveness matrix, then an assignment that minimizes the total effectiveness in one matrix also minimizes the total effectiveness of the other matrix."

The next step in the procedure is to subtract the minimum element in each row from all the elements in its row, giving:

	I	II	III	IV
A	0	18	9	3
B	9	24	0	22
C	23	4	3	0
D	9	16	14	0

THE EFFECTIVENESS MATRIX

TABLE 13

Then we subtract the minimum element in each column from all the elements in its column, resulting:

	I	II	III	IV
A	Ø	14	9	3
B	9	20	Ø	22
C	23	Ø	3	0
D	9	12	14	Ø

THE EFFECTIVENESS MATRIX

TABLE 14

As long as our matrix consists of positive or zero elements, the total effectiveness cannot be negative for any assignment. It now becomes obvious that if we can select an assignment that has a zero total, there cannot be an assignment with a lower total. This simply means that the total has to be minimum if all assignments can be made to positions where there are zero elements. On the basis of the above matrix, the optimum assignment will be:

<u>A-I, B-III, C-II, D-IV</u>

For clarity and simplicity we have used an example which provided us with an obvious solution of the problem after reduction of the Effectiveness Matrix by subtraction. But there will be other cases where a complete assignment may not exist among the zeros, or even if it exists, it may be difficult to identify the complete assignment if the matrix is of large dimensionality. Thus we have to have algorithms for finding the maximal existing assignment among the zeros of a matrix with some zeros and non-negative remaining elements, and for obtaining more zeros by further modifying a matrix by additions or subtractions to rows or columns when a complete assignment does not exist among the zeros. In all cases the following rules are used to start with:

(1) Examine rows successively until a row with exactly one unmarked zero is found. Mark (□) this zero, as an assignment will be made there. Mark (X) all other zeros in the same <u>column</u> to show that they cannot be used to make other assignments. Proceed in this fashion until all rows have been examined.

(2) Next examine columns for single unmarked zeros, marking them (□) and also marking with an (X) any other unmarked zeros in their rows.

(3) Repeat (1) and (2) successively until one of two things occurs: a) there are no zeros left unmarked, or b) the remaining unmarked zeros lie at least two in each row and column.

In outcome (a) we have a maximal assignment. In outcome (b) we must use ingenuity and/or trial and error in order to build up to a maximal assignment so that we may avoid using a highly complex algorithm to keep the methodology simple and practical. If by the application of the above rules, we can obtain the maximal assignment with an assignment in every row, this maximal assignment is a complete solution to the original problem. However, if it does not contain an assignment in every row, we have to modify the effectiveness matrix by addition or subtraction. Before going into that problem, we will work out an example of finding maximal assignments.

Following is a matrix with zero elements in the positions shown, and positive non-zero elements elsewhere. Our problem is to find a maximal assignment.

	0			
0	0			
0		0		0
0		0		0
		0	0	0

THE EFFECTIVENESS MATRIX

TABLE 15

By following the rules, we find that row 1 has a single zero in column 2. We make an assignment there and delete the second zero in column 2.

	[0]			
0	X			
0		0		0
0		0		0
		0	0	0

THE EFFECTIVENESS MATRIX

TABLE 16

Row 2 has a single zero in the first column. We make an assignment there and delete the remaining zeros in column 1.

	[0]			
[0]	X			
X		0		0
X		0		0
		0	0	0

THE EFFECTIVENESS MATRIX

TABLE 17

All the remaining rows have at least two zeros left; so we now examine columns. Column 4 has a single zero in row 5; so we make an assignment there and delete the remaining zeros in row 5.

	[0]			
[0]	X			
X		0		0
X		0		0
		X	[0]	X

THE EFFECTIVENESS MATRIX

TABLE 18

Both the remaining rows and columns have two zeros. We make an assignment in the position (3,3) and delete the remaining zeros in row 3 and column 3. This leaves one zero at (4,5) and we make the last assignment.

	[0]			
[0]	X			
X		[0]		X
X		X		[0]
		X	[0]	X

THE EFFECTIVENESS MATRIX

TABLE 19

There are no remaining zeros, as we can see, and every row has an assignment. Since no two assignments are in the same column, the maximal assignment is a solution to our problem. However, we have to remember that there may be more than one maximal assignment.

Now we turn to the remaining case where the maximal assignment does not give us a complete assignment. How should we add further zeros? The following rules and their repeated application will lead to a complete optimal assignment in a finite number of iterations:

Starting with a maximal assignment:

(1) Mark all rows for which assignments have not been made.

(2) Mark columns not already marked which have zeros in marked rows.

(3) Mark rows not already marked which have assignments in marked columns.

(4) Repeat steps (2) and (3) until the chain of markings end.

(5) Draw lines through all unmarked rows and through all marked columns. There should be as many lines as there were assignments in the maximal assignment, and every zero will have at least one line through it. This method yields the minimum number of lines that will pass through all zeros.

(6) Having drawn the set of lines in steps (1) through (5), examine the elements that do not have a line through them. Select the smallest of these, and subtract it from all the elements that do not have a line through them. Add this smallest element to every element that lies at the intersection of two lines. Leave the remaining elements of the matrix unchanged.

To illustrate, now we construct the minimum number of lines that will pass through all the zeros of the matrix below.

	0			
0			0	
		0	0	0
0				
			0	

THE EFFECTIVENESS MATRIX
TABLE 20

We first mark the maximal assignment.

	[0]				
X			X		✓
		[0]	X	X	
[0]					✓
			[0]		✓
✓			✓		

THE EFFECTIVENESS MATRIX
TABLE 21

Then we mark row 2 as having no assignment and columns 1 and 4 as having zeros in row 2. Next mark rows 4 and 5 because they contain assignments in marked columns. The procedure leads to no further marked rows or marked columns. The minimum set of lines that will cover all zeros is the set through rows 1 and 3 (unmarked) and columns 1 and 4 (marked).

Now we modify the matrix below so as to obtain a better maximal assignment:

5	0	8	10	11
0	5	15	0	3
8	5	0	0	0
0	6	4	2	7
3	5	6	0	8

THE EFFECTIVENESS MATRIX

TABLE 22

The zeros are in the same position as in the previous example; so we already have the maximal assignment and the lines as shown below.

5	0	8	10	11
0	6	15	0	3
8	5	0	0	0
0	6	4	2	7
3	5	6	0	8

THE EFFECTIVENESS MATRIX

TABLE 23

Now we select the smallest element not deleted by a line; in this matrix it is 3 in row 2, column 5; we subtract this element from every element that does not have a line through it, and add it to every element that lies at an intersection of two lines. The new matrix is the following:

8	[0]	8	13	11
~~0~~	3	12	~~0~~	[0]
11	5	[0]	3	~~0~~
[0]	3	1	2	4
3	2	3	[0]	5

THE EFFECTIVENESS MATRIX

TABLE 24

We now find that we have a complete assignment in positions with zero elements (1,2); (2,5); (3,3); (4,1); (5,4). If the maximal assignment did not constitute a solution to the original problem, we would proceed to draw lines and continue to iterate until we finally obtained a solution.

XIII. THE SEQUENCING MODEL[1]

In an information system, in the chain of input - processing - output, there will be problems of sequencing that may be adequately handled by adaptation of the techniques used in the job shop. In sequencing we are concerned with a situation where the effectiveness measure is a function of the order or sequence in which a series of tasks are performed. Information systems receive input in packages of different content and format like books, periodicals, R & D Reports, manufacturing information, marketing and financial information, etc., in macro- and microforms. These need different treatment on different equipments in different order or sequence.

These problems may be categorized under two groups. In the first group, we have n tasks to perform, each of which requires processing on some or all of m different equipments. The effectiveness of any given sequence of the tasks at each equipment can be measured, and we would like to select from the $(n!)^m$ theoretically possible sequences or orders, one (or several) which optimizes the effectiveness measure, out of those which satisfy the restrictions on the order or sequence in which each task must be processed through the m equipments. Theoretically, solution by enumeration is always possible, but the likely number of cases for enumeration make this approach impractical even for moderate values of m and n.

We have in the second group a number of equipments and a set of tasks to perform. We have to decide on the next task to be started on an equipment that has just completed a task, keeping in mind that the

[1] Ibid., pp. 250-258, passim.

set of tasks is liable to grow unpredictably with time. Solutions are known only for some special cases of the first group and there appears to be no mathematical technique for solution of the second group of problems. Following are some specific case illustrations of processing each of n tasks through m equipments.

There are n tasks (1, 2, ..., n), each of which has to be processed one at a time at each of m equipments. The order of processing each task through the equipments is given (for example, task 1 is processed at equipments A, C, B, in that order). We assume that we know the exact time each task must spend at each equipment. The problem is to find a sequence for processing the tasks so that the total elapsed time for all the tasks will be at a minimum.

Symbolically, let

A_i = time for task i on equipment A;

B_i = time for task i on equipment B, etc.;

T = time from start of first task to completion of the last task.

We wish to determine for each equipment a sequence, $(i_1, i_2, \ldots, i_n)$, where $(i_1, i_2, \ldots, i_n)$ is a permutation of the integers (1, 2, ..., n), which will minimize T.

Following are the three special cases for which satisfactory mathematical solutions are available:

(1) n tasks and two equipments A and B; all tasks processed in the order AB;

(2) n tasks and three equipment A, B, and C; all jobs processed in the order $A \cdot B\ C$; other limitations given with the illustration;

(3) two tasks and m equipments; each task to be processed through the equipments in a prescribed order which is not necessarily the same for both tasks.

Following is an illustration of processing n tasks through two equipments, for which a solution is available:

(1) Only two equipments are involved, A and B;

(2) Each task is processed in the order A B;

(3) The exact or expected processing times A_1, A_2, ..., A_n, B_1, B_2, ..., B_n are known.

The problem is to minimize T, the elapsed time from the start of the first task to the completion of the last task. The following method of computation for the solution of the problem is due to Johnson.[2]

(1) Select the smallest processing time occurring in the list A_1 ... A_n, B_1 ... B_n. If there is a tie, select either smallest processing time.

(2) If the minimum processing time is A_r, do the rth job first. If it is B_s, do the sth last. This decision will apply to both equipments A and B.

(3) There are now n-1 tasks left to be ordered. Apply steps 1 and 2 to the reduced set of processing times obtained by deleting the two equipment processing times corresponding to the task already assigned.

(4) Continue in this manner until all jobs have been ordered. The resulting ordering will minimize T.

---

[2] S. M. Johnson, "Optimal Two- and Three-Stage Production Schedules with Setup Times Included," Naval Research Logistics Quarterly, I, No. 1(March, 1954), pp. 61-68.

To illustrate, we have five tasks, each of which must go through the two equipments A and B in the order A B. Processing times are given in Table 25, below.

Task	PROCESSING TIME, HR. Equipment A	Equipment B
1	5	2
2	1	6
3	9	7
4	3	8
5	10	4

TABLE 25

We have to determine a sequence for the five jobs that will minimize the time T. Applying the method above, we find that the smallest processing time is 1 hour for task 2 on equipment A. Thus we schedule task 2 first:

2				

The reduced set of processing times is:

Task	A	B
1	5	2
3	9	7
4	3	8
5	10	4

The smallest processing time, 2, is B_1. So, according to the method we schedule task 1 last:

2				1

Continuing in the same manner, the next reduced set of processing time we have is:

Task	A	B
3	9	7
4	3	8
5	10	4

This will give us the schedule:

2	4			1

leaving the remaining set of processing time:

Task	A	B
3	9	7
5	10	4

This will give us the schedule:

2	4		5	1

So that the optimal sequence is:

2	4	3	5	1

Elapsed time corresponding to the optimal ordering can be calculated now using the individual processing times given in the statement of the problem, as shown below:

Task	Equipment A Time in	Equipment A Time out	Equipment B Time in	Equipment B Time out
2	0	1	1	7
4	1	4	7	15
3	4	13	15	22
5	13	23	23	27
1	23	28	28	30

Thus the minimum elapsed time is 30 hours. Idle time is 3 hours for Equipment B, and 2 hours for Equipment A.

Now we try an example of processing n tasks through three equipments. At present no method is available for the solution of this problem of sequencing n tasks, three equipments, A, B, and C, prescribed order A B C for task and no passing. However, the method of sequencing n tasks through two equipments, as described above, can be extended to cover the special cases where either or both of the following conditions hold:

(1) The smallest processing time for equipment A is at least as great as the largest processing time for equipment B.

(2) The smallest processing time for equipment C is at least as great as the largest processing time for equipment B.

The method is to replace the problem with an equivalent problem involving n tasks and two equipments. The two fictitious equipments are denoted by G and H, and the corresponding processing times G_i and H_i are defined by:

$$G_i = A_i + B_i$$

$$H_i = B_i + C_i$$

The problem is worked out with prescribed ordering G H, according to the previous method. Let us have five tasks, each of which must go through the equipments A, B, and C in the order of A B C. Processing times are:

Task	A	B	C
1	4	5	8
2	9	6	10
3	8	2	6
4	6	3	7
5	5	4	11

Our problem is to determine a sequence for the five tasks that will minimize the elapsed time T. Here we have min A_i = 4, max B_i = 6, min C_i = 6. Since max $B_i \leq$ min C_i, we are justified in applying the previous method. The equivalent problem becomes:

Task	G	H
1	9	13
2	15	16
3	10	8
4	9	10
5	9	15

Because of the ties, there are several optimal orderings. They are:

1	4	5	2	3

1	5	4	2	3

4	5	1	2	3

4	1	5	2	3

5	1	4	2	3

5	4	1	2	3

Any of these orderings may be used to sequence the tasks through equipments A, B, and C; and they will all yield a minimum elapsed time of 51 hours.

XIV. DISCUSSION

The previous section demonstrates the application of the information system design methodology that has been developed in this dissertation. We started with a hypothetical information system, after redefining the "modified PERT activity." The process goes through the identification of the activities involved and organization of these activities in an Umbrella Network. The next step in the process was to develop a PERT network of the Subject Indexing Function of MEDLARS and indicate the activity in the Umbrella Network that subsumed this function. Having done this, we zeroed in on one of the activities of the MEDLARS indexing function network. This was the activity 8-9 (Indexing) in the network.

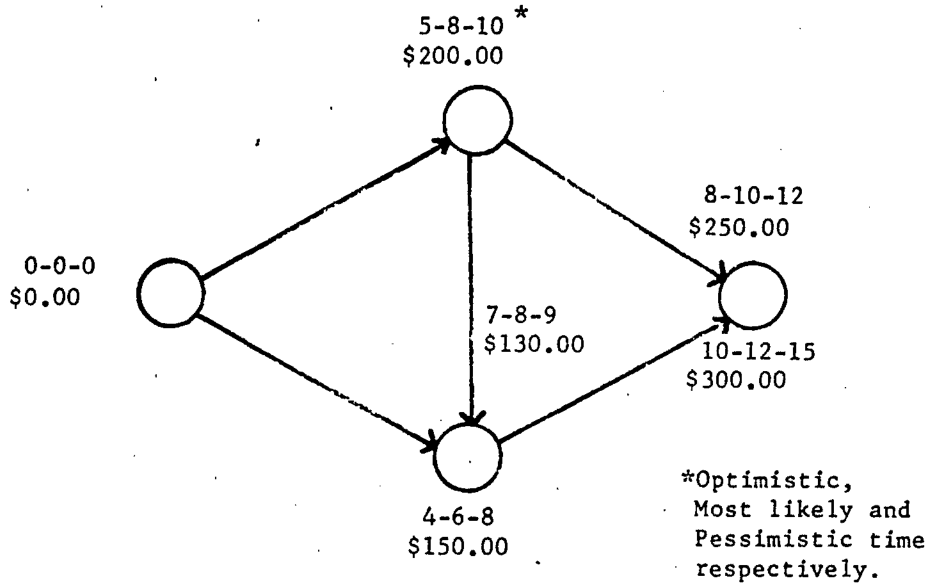
The process then asked for taking this activity 8-9 through the PERT Computational Program, CPM Computational Procedure, the Assignment Model, and the Sequencing Model. However, since an activity is no different than any other in the network, so far as the treatment it receives as a modified PERT activity, it was preferred to demonstrate the application of the PERT Program, CPM Procedure, Assignment, and Sequencing in a general way, notwithstanding the fact that the activity 8-9 could be any one of the activities used in the demonstration.

As has been stated before, the PERT Program, the CPM Procedure, and the Assignment and Sequencing Models are in fact control mechanisms embedded in the basic functional unit. They control time, cost, assignment, and sequence of each activity in the network, and they will provide for the "continuous system monitoring" at the basic functional unit level so far as time, cost, assignment, and sequencing are concerned.

Having said all this, we may now look back to the design methodology that has been developed and see how it can help the designer to design information systems with a built-in control mechanism at the basic functional unit level. We have seen that it is not enough to have control stated as a system objective (vide p. 66 , No.8). Control should be specified as a design requirement to assure its existence in the system at the level where the basic functional activities are taking place. Can our design methodology accomplish this? Can our methodology help develop design requirements from the diagnostics generated by the system operating experience and create design algorithms which will force the designer to go through the process of problem solving at the point of their logical occurrence on the drawing board? Our answer is--yes it can. So, let us see how.

Our methodology is based on networking technique. The nodes of the network are the components of the system. The process starts with an Umbrella Network. Then by the application of the family networking technique, the Umbrella Network is gradually unfolded until it reaches a level of specificity where the basic functional units (nodes or activities) are identified. At each level, the nodes are networked in a precedence and dependency relationship; each node is fixed in the network in a logical interrelationship. But this is not rigid or irrevocable. Both the identity of the nodes and their interrelationships with respect to each other may change. The design in the form of the network is never frozen. It moves in parallel with the design, implementation, and operation, in the form of a graphical representation of the physical system, allowing for manipulation of interrelationship at any time.

Now since our methodology is based on PERT/CPM technique, each of the activities in the network or subnetwork, as the case may be, will have associated time and cost data like the following Figure 31.

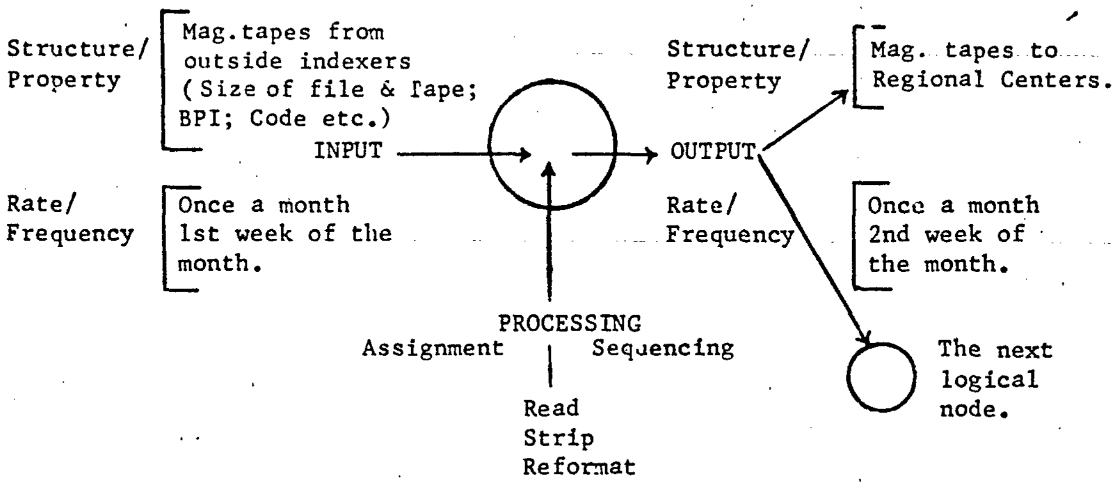


DISPLAY OF TIME AND COST DATA

Figure 31

The initial raw data will come from the records of the system's operating experience, or in its absence, from the educated guesses of the experienced technical personnel. The PERT Computational Program and the CPM Computational Procedure of the methodology will compute the time and cost estimates respectively. This process will generate the critical path through the network or subnetwork and optimize the cost. The information that will be generated by this process will help the management to manage by "exception" by drawing the management's attention only to those critical activities which need to be "crashed" and to those "slack" activities from where resources may be diverted. This will optimize the resource allocation problem, along with the control of time and cost at the basic functional unit level.

But we have also modified PERT/CPM in our methodology. The methodology demands that for each activity in the network, the input, processing, and output be specified. So, for each activity, the design must provide the structure, property, rate, and frequency of both input and output. The design must also provide for each activity the manner of processing the input to generate the output which becomes the input to the next logical activity, as illustrated in the following Figure 32.



INPUT-PROCESSING-OUTPUT FLOW
Figure 32

In other words, the design must specify and optimize the assignment of jobs to man-machine capabilities for each activity at the basic functional unit level. The Assignment Model and the Sequencing Model of the methodology will provide algorithms to optimize and control assignment and sequencing of jobs. The process will also generate information which will be recorded and used as input for the next scheduling of the operation of the system. The methodology does not provide algorithms for handling the structure, property, rate, and frequency of the input and output of the activities, but points to the relevant literature for possible solutions.

Thus we see that the information system design methodology that has been developed can manipulate the activity interrelationship in a continuous manner, control and optimize the activity time, cost, assignment, and sequencing, and provide the system the capability of "continuous system monitoring," at the basic functional unit level.

An operational example of how control is built into a basic functional unit would be in order at this point. We have chosen as an illustration the activity 8-9 (indexing) of the Subject Indexing Function of MEDLARS (PERT Network p.131), with specific reference to the control of assignment of documents to indexers.

Let us see what is involved in the process. Indexing is the process of identification of the intent of the authors as expressed in the documents to be indexed: What is the document all about? What does the author want to prove or demonstrate? What are the questions that this document deals with or resolves? Who are the people to whom this document is addressed? These are some of the questions that the indexer normally asks himself when indexing a document. The indexer then tries to translate the answers to these and similar questions into the vocabulary or terminology that is legal in the system, or into the terminology that he thinks adequately reflects the concepts in the document, in case the vocabulary is not controlled.

This is a subjective task liable to suffer from interindexer inconsistency and will remain so until the perfection of automatic indexing. The indexers will vary in their efficiency, depth of subject knowledge, language proficiency, and sensitivity to the system and user requirements. On the other hand, the documents will vary in subject matter, sophistication of treatment, language, usefulness to the system, and so forth.

So the in-charge of the indexing activity has the problem of matching the indexing tasks with the available indexing capabilities. If the in-charge had a unique indexing capability available for each indexing task, then there would be no assignment problem. But because this is not the case, most of the time, the in-charge will have an assignment problem. The problem is, how should the job be assigned one to an indexer, so as to minimize the total man-hours.

From the previous performance records, the in-charge will know the individual proficiencies of the indexers. These records will be maintained by the in-charge and will be used whenever an assignment has to be made. (The indexer performance record may take the following form, Figure 33.)

Indexer	Subject Knowledge Major	Minor	Language Proficiency H (High) M (Medium) L (Low)	Average speed of Indexing Min./Doc. Major	Minor	Comments (Sensitivity to system & user requirements, etc.)	Q of Waiting docs.
Winsky	Artificial Intelligence Computer Simulation Heuristics	Psychology Teleonomy Educational Measurement	Russian (H) Polish (L)	8	15	System (H) User (L)	9
Klein	Electronic Engineering Logic Circuitry	Analog Computers	None	15	26	System (H) User (L)	5
Smith	Industrial Engineering Statistics	Management Psychology	Spanish (H)	10	17	System (L) User (H)	8
Harris	Programming Languages Computational Linguistics	Linguistics	Russian (H) German (L)	8	11	System (H) User (H)	10

INDEXER PERFORMANCE RECORD

Figure 33

This has been done in our Assignment Model (The Effectiveness Matrix, p. 180, repeated here to facilitate reference). For example, row A has the numbers 8, 26, 17, and 11 on columns I, II, III, and IV, respectively. These numbers could be minutes, hours, or days, and they represent the estimates of the in-charge as to how long a certain indexer would take if a particular document is assigned to him for indexing.

ANALYSTS

DOCUMENTS	I	II	III	IV
A	8	26	17	11
B	13	28	4	26
C	38	19	18	15
D	19	26	24	10

THE EFFECTIVENESS MATRIX

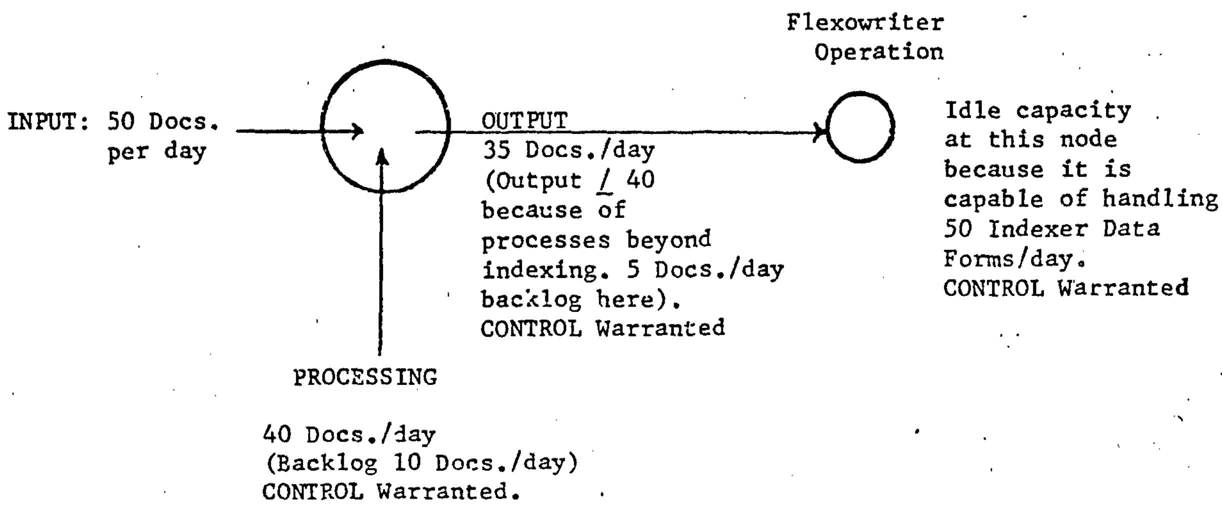
TABLE 11

In our Assignment Model, it can be seen that Analyst II (or indexer) would take, perhaps, 26 minutes if document A is assigned to him for indexing, whereas Analyst I would take only 8 minutes if the same document is assigned to him, and so forth. It is assumed that the quality of indexing would remain the same, that is, Analyst I is not indexing document A in shorter time because he is doing a quick and dirty job. Thus we see in our Assignment Model how the in-charge minimizes the indexing time and achieves an optimum allocation of indexers to documents, and thereby establishes control over the indexing function.

Because the in-charge maintains indexer performance and proficiency records, he will be in a position to estimate how many documents of what nature his staff can handle, say, per day. His indexer performance record will provide him with the necessary information for the computations.

He will also know, because of the availability of the records, how many times in a day or how frequently he can receive an input from the predecessor activity and produce an output for the successor activity.

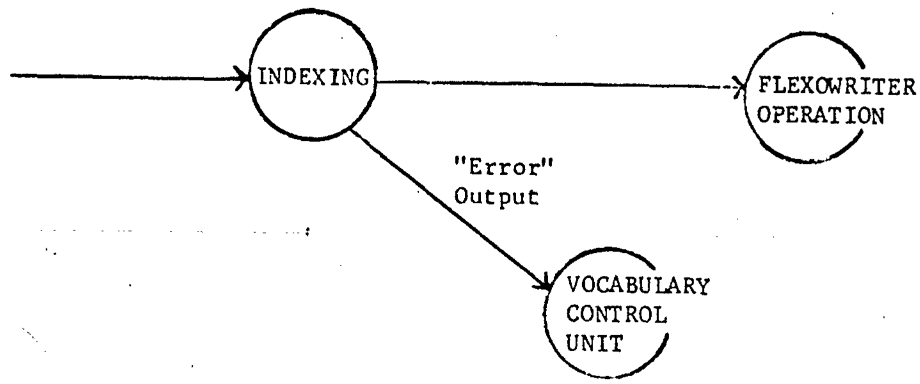
So any bottleneck at the input/output ligands or any backlog inside the activity of indexing itself would be noticeable immediately and even be predictable, if the current rate is projected. Therefore, the in-charge can take some control action with respect to the predecessor or successor activity. On the other hand, the process may make the in-charge aware of any idle capacity and help him reallocate it to some other activity in the network, as illustrated in the following Figure 34.



DETECTION OF IDLE CAPACITY

Figure 34

The indexer may find it difficult to index some documents due to the inadequacies in the vocabulary if the vocabulary is controlled. In that case, there will be some "error" output which will be input to the vocabulary control activity where the necessary corrective measure will be taken, as shown in Figure 35.



"ERROR" OUTPUT

Figure 35

Thus we see that the whole process works like a chain reaction. An activity is an integral part of a network, influencing and being influenced by the other activities in the network. Each activity has the capability of controlling time, cost, assignment, and sequencing through the application of PERT Program, CPM Procedure, Assignment Model, and the Sequencing Model. The network, which graphically represents the different phases of the physical system--design, implementation, and operation at different points in time--at all times moving in parallel with the system, is optimized because each individual interactive and interdependent activity making up the network is optimized, so far as the time, cost, assignment and sequencing are concerned.

XV. CONCLUSIONS

This dissertation attempts to demonstrate that PERT/CPM, or some modified version thereof, can be developed into an information system design methodology. PERT/CPM is a networking technique with time estimation and cost computation capabilities for project control and optimum allocation of resources. All systems, including information systems, are composed of interacting and interdependent components. A network of these components establishes the physical system. The network links indicate the functional flow of the system. We have seen that such network representation of an information system can be accomplished with PERT/CPM. We can also have the system represented at different levels of generality and specificity, so we can get down to the basic functional unit level of the system

When we have done this we face a different kind of problem. How do we know that the system is going to work? How can we make certain that the system will live through its life expectancy and perform its design function? For this we looked into the basic functional unit of the system -- the activity. Every activity in the network will receive an input from the predecessor activity, process the input, and produce an output which becomes an input to the successor activity. If we could establish control over these activities or, in other words, if we could establish control over the internal processes and external interrelationships of the basic functional units of the network, we would be able to assure system survival and optimize system performance.

This has been done by the modification of the PERT activity. The PERT activity has been redefined to include the input, processing, and output elements. Each activity will study its input and output to determine its structure, property, rate, and frequency. Processing transforms the input. This job of transformation has to be assigned and sequenced through the man-machine capabilities of the basic functional unit. This is done by the application of the Assignment, and Sequencing Models. Thus control over processing is established.

In his MEDLARS Evaluation Study, Lancaster concluded that "continuous system monitoring" is ultimately essential to the success of any large retrieval system. "A single evaluation study, however comprehensive, cannot be expected to discover more than a very small fraction of the specific inadequacies of the system."

An office or staff assigned to monitor quality, cannot do the job because it will have no direct involvement in the continuous operations of the basic functional units, and its actions will have to wait until something that warrants control action surfaces, overcoming the "gravitational pull" of the system hierarchy. Unless monitoring is continuous, as Calvin Mooers (p.114), Saul Herner (p.115), and Glaser[1] also pointed out, and unless it is incorporated into the basic functional units of the system by making it a design requirement, findings of the monitoring office or one-time evaluations, will be contextually irrelevant to the system. In other words, monitoring of a large information system, sometimes with geographically dispersed subsystems, can only be continuous if control is incorporated in the basic functional units where the activities are taking place.

---

[1]R. Glaser and D. J. Klaus, "Proficiency Measurement: Assessing Human Performance," in Psychological Principles in System Development, ed. by R. M. Gagne (New York: Holt, Rinehart, and Winston, 1962), pp. 419-474.

So there is a need for an information system design methodology which can handle the problem of incorporating control in the basic functional units which are ultimately networked into the desired system.

The information system design methodology that has been developed here fulfills this need. By providing the means to control the time, cost, assignment, and sequencing of the activities of the basic functional units and a way to network them into the desired system, the methodology will help the designer to create adaptive information systems. In other words, the designer, the management, and the system operators will have a methodology for optimum resource allocation, time-scheduling, optimizing system performance, and continuous system monitoring.

XVI. POSSIBLE AREAS OF RELATED RESEARCH

The possible areas of related research have been indicated in the appropriate places of the text as they occurred. While developing the information system design methodology, it has been felt that the general areas of systems theory, control theory, and operations research have a great deal to offer towards the sophisitication and quantification of information system design, implementation, operation, and evaluation. Philip Morse's "Library Effectiveness" is a giant step to this end.[1]

Like most systems, information handling systems face the problem of crowding, congestion, and bottleneck. Application of queuing theory may alleviate many of these problems.

Prevention is known to be better than cure. Most often it is possible to prevent if we can predict more or less accurately. We have seen that MEDLARS had to rely on the outside indexers, due to the unforeseen growth of the input, though in-house indexing was the original philosophy. Experimental application of probability theory in general and Markov process in particular may give us insight in the area of prediction. Libraries and information systems have never had to justify their existence by showing a profit. But information is fast becoming a commodity and sooner or later will have to submit to the economics of price theory.

The isomorphism of information processing in the artificial and the biological system is intriguing. The importance of the study of

---

[1]Phillip M. Morse, Library Effectiveness: A Systems Approach (Cambridge: The MIT Press, 1968).

information processing in the biological system in general and neural network, genetic code, memory, learning, forgetting, central and peripheral nervous system in particular, to determine their possible application in the library and information system design and operation, can hardly be overemphasized.

An immediate area of application is obviously the design, development, and operation of an information system by applying the methodology developed in this dissertation. This work also can serve as a basis for the development of a course on information system design methodology.

REFERENCES

Ackoff, Russell L., and Sasieni, Maurice W. Fundamentals of Operations Research. New York: John Wiley & Sons, Inc., 1968.

Archibald, Russel D. PERT Management Information Systems. Culver City, Calif.: Hughes Aircraft Corporation, 1962.

Austin, Charles J. MEDLARS 1963-1967. Public Health Service Publication No. 1823. Bethesda: National Library of Medicine, 1969.

Baker, Bruce N., and Eris, Rene L. An Introduction to PERT-CPM. Homewood, Ill.: Richard D. Irwin, Inc., 1964.

Churchman, C. W. Does Operations Research=Systems Science? Symposium on Operations Research. Santa Monica, Calif.: System Development Corporation, March 27, 1963.

Cleland, David I., and King, William R. Systems Analysis and Project-Management. New York: McGraw-Hill Book Company, 1968.

DoD and NASA Guide, PERT Cost Systems Design. Washington, D.C.: Office of the Secretary of Defense, June 1962.

Evarts, Harry F. Introduction to PERT. Boston: Allyn and Bacon., Inc., 1964.

Farradane, J. "Information for Design." The Design Method. Edited by S. A. Gregory. New York: Plenum Press, 1966.

Glaser, E. L., and Klaus, D. J. "Proficiency Measurement: Assessing Human Performance." Psychological Principles in System Development. Edited by R. M. Gagne. New York: Holt, Rinehart, and Winston, 1962.

Goode, H. H., and Machol, R. E. Systems Engineering: An Introduction to the Design of Large-Scale Systems. New York: McGraw-Hill Book Company, 1957.

Hall, A. D. A Methodology for Systems Engineering. Princeton, N. J.: Van Nostrand, 1962.

Healy, Thomas L. Project Administration Techniques. Dayton, Ohio: The National Cash Register Co., April 1, 1963.

Herner, Saul. System Design, Evaluation and Costing--In Plain English. Contract No. AF49(638)-1424, Project No. 9769-01. Washington, D. C.: Herner and Company, 1969.

Holzman, A. G. "Critcal Path Methods." Encyclopedia of Library and Information Science. Edited by Allen Kent and Harold Lancour. Vol. V. New York: Marcel Dekker, Inc., in press.

Jaffe, J. "The System Design Phase." Developing Computer-Based Information Systems. By Perry E. Rosove. New York: John Wiley & Sons, Inc., 1967.

Johnson, S. M. "Optimal Two- and Three-Stage Production Schedules with Setup Times Included." Naval Research Logistics Quarterly, I, No. 1 (March, 1954), 61-68.

Kent, Allen. Textbook on Mechanized Information Retrieval. 2nd Ed. New York: Interscience Publishers, 1966.

Kent, Allen. Specialized Information Centers. Washington, D. C.: Spartan Books, 1965.

Kent, A.; Belzer, J.; Kurfeerst, M.; Dym, E. D.; Shirey, D. L.; and Bose, A. "Relevance Predictability in Information Retrieval Systems." Method. Inform. Med., VI, No. 2(April, 1967),45-51.

King, W. R., and Wilson, T. A. "Subjective Time Estimates in Critical Path Planning: A Preliminary Analysis." Management Science, XIII, No. 5 (January, 1967), 307-320.

Lancaster, F. Wilfrid. Evaluation of the MEDLARS Demand Search Service. Washington, D. C.: U. S. Department of Health, Education, and Welfare, January, 1968.

Lancaster, F. Wilfrid. Information Retrieval Systems; Characteristics, Testing, and Evaluation. New York: John Wiley & Sons, Inc., 1968.

Malcolm, D. G.; Rosebloom, J. H.; and Frazer, W. "Application of a Technique for Research and Development Program Evaluation." Operations Research, VII, No. 5 (Sept.-Oct., 1959), 646-669

Marshall, A. W., and Meckling, W. H. Predictability of the Costs, Time and Success of Development. Paper P-1821. Santa Monica, Calif.: Rand Corp., December 11, 1956.

Methodologies for System Design. Final Report on Contract No. AF 30(602)-2620, Project No. 4594, Task No. 459-405. Los Angeles: Hughes Dynamics, 1964.

Morse, Philip M. Library Effectiveness: A Systems Approach. Cambridge: The MIT Press, 1968.

The National Library of Medicine. Annual Report for the Fiscal Year 1968. Washington, D. C.: Government Printing Office, 1969.

Rosove, Perry E. Developing Computer-Based Information Systems. New York: John Wiley & Sons, Inc., 1967.

Sasiene, Maurice; Yaspan, Arthur; and Friedman, Lawrence. Operations Research--Methods and Problems. Wiley International Edition. New York: John Wiley & Sons, Inc., 1959.

Scharf, Tom. "Management and the New Software." Datamation, XIV, No. 4 (April, 1968), pp. 52, 57, 59.

Watts, Ronald D. "The Elements of Design." The Design Method. Edited by S. A. Gregory. New York: Plenum Press, 1966.

Zalokar, Frederick J. The Critical Path Method; A Presentation and Evaluation. Schenectady, New York: General Electric, May 18, 1964.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) The University of Pittsburgh Pittsburgh, Pennsylvania 15213	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP --

3. REPORT TITLE

INFORMATION SYSTEM DESIGN METHODOLOGY BASED ON PERT/CPM NETWORKING AND OPTIMIZATION TECHNIQUES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report 4/15/68 - 8/1/70

5. AUTHOR(S) (First name, middle initial, last name)

Anindya Bose

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1970	214	33

8a. CONTRACT OR GRANT NO. N00014-67-A-0402-0004 b. PROJECT NO. ONR:437:MD:FS c. NR.049-258 d.	9a. ORIGINATOR'S REPORT NUMBER(S) Nooo14-67-A-0402-0004 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) --

10. DISTRIBUTION STATEMENT

Distribution of this Document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
NONE	Office of Naval Research Department of the Navy Washington, D.C. 20360

13. ABSTRACT

This dissertation attempts to demonstrate that the Program Evaluation and review Technique (PERT)/Critical Path Method (CPM) of some modified version thereof, can be developed into an information system design methodology. The need for such a methodology has been identified through a study of the characteristics and problems of information systems in general, and Lancaster's Evaluation of the MEDLARS Demand Search Service (January 1968) in particular, which concluded that "continuous system monitoring is ultimately essential to the success of any large retrieval system."

The methodology utilizes PERT/CPM which isolates the basic functional units of a system and sets them in a dynamic time/cost, precedence and dependency inter-relationship network. To monitor the internal activity of the basic functional units, the methodology applies Assignment and Sequencing algorithms. By providing the means to control the time, cost, assignment, and sequencing of the activities of the basic functional units in a continuous manner, and a way to network them into the desired system, the Methodology fulfills the need for a continuous monitoring information system design methodology. An interactive mode PERT Computer Program has been written in PIL/L which runs on the IBM 360/50 Pitt Time Sharing System using the 2741 Terminal.